

RA Fraser C.M.; Barrell B.; human malaria parasite Plasmodium
 RT "Genome sequence of the human malaria parasite Plasmodium
 RT falciparum.";
 RL Nature 415:498-511 (2002).
 DR EMBL; AR014841; ANN5938.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 227 AA; 262841 MW; 6D5D4BRRR643339C CRC64;

Query Match 6.3%; Score 252.5; DB 2; Length 227;
 Best Local Similarity 19.1%; Pred. No. 0.024; Mismatches 319; Indels 323; Gaps 49;
 Matches 190; Conservative 155; Mismatches 319; Indels 323; Gaps 49;

QY 10 FPKLNGKE-----GSILKQDTGVER-----HHDETEESKEK 42
 Db 690 YNNNNNNKETCTSINIKHSENKYPFNKHSHVNSYKNTNLPHRNALTTSNNRNNEVEREK 749

QY 43 SFTFLRNIST----IRDPEK---DLKLU-----KKEERKVDFT----SB 79
 Db 750 EK---DRNITGNNNNVVEYNNSCIPPLKQKIPDGEVRNSINKLNANVNTORTSSVSY 806

QY 80 TCKRMBEYDY-----KDDKAMIAVDDGTDLYETEKUDRIKSYKYGVLSPSK 128
 Db 807 TNNKLDNSPMPILINGIRBSCKLNNNL--NGNSIGFNSKLDVYHHSMV----- 858

QY 129 DCHPBLIGKISVSKNQKVVYGRNYSKEIKATKTYDFHSKTMF-----DLYANIN-- 179
 Db 859 -NESYPLKMMGRN-----IEHNDKNNIFLVKMYETTSNIHNGI 900

QY 180 -----DITVGLAFAGDR-----LFVKDN---DOKK----- 202
 Db 901 HENSMILKNNYLNKACTPHGYSRHNQKOMYTBEMLNINQKKNYHYHNGTVLKLPLVNTNN 960

QY 203 -----AEIKRMPKKEKTS--EVYVYSSGNGVIEGLGDLSKNRPDILTRMESGKI 253
 Db 961 VAVNBRADINNSAQKRLHSLSKMGYBOKSMRMEYRNKLYNNINNNNNNN-----NTI 1013

QY 254 YSDSEKQOY--LLKONILRKGYAL-----KVTTAEGKTDLEGNGVSYKEDIAK-- 302
 Db 1014 YDNDEYQYNNSYCFHDSDKOMPFLHQNISKLLTHSNNKNSKNSPENGINVESEGHHLANPEI 1073

QY 303 -----TOKANP-----N 309

Db 1074 KTFARNSYPIALNQGLINCPLQCLGTDQNQRMKHNWVYIKRNEYLNNKIGTINVLKREG 1133

QY 310 IRLS-----ETTYADSAPNVRDERSRSTOSVLSMALDGENILIRYQVFTFND 356
 Db 1134 LAKISTHNGKESFSNMDNTRNTYMEGANIQDVAWNNNKESC-DNTKMRTRISLNFTSRE 1192

QY 357 K-GB--AIDKQGNLVTDSKULVFGK-DDKETGKEDXENRAKED-GSMC-RID---TKP 408
 Db 1193 SYGERHESLDVYQECYVNNKLU--NRKNDKCY--EDMNNTYNTEDDNASMKQFYEFTNSNP 1248

QY 409 VNLSDM---KVNPNPSKSNKNTVNRNPEFYLRGKLSDKGFWME----LRYNESV--- 455
 Db 1249 YIVDQERINMKVWN---NVLJDNNSWYVUDSKYDSEKNAENKSDILANNENIHLKD 1304

QY 456 -----VDNYLIYGDLEHIDNTDPTKLNQDGDMW--GMDYKANGPPDKV-TMDG 506
 Db 1305 OKKKKIQNNRERFISQADINRNSOBEVYERKHEPL-WVIMASNEERKSYISLYSDMSS 1363

QY 507 NYLQTCYSDPLAKAKVWVHYPOLYDNUKPEVNIDPKGNTSTEVADGKSVWVNINDRNG 566
 Db 1364 KRTVKTKYSDMNNVNEVILANBNDLMLTTERKVKVQLEGENKNDMAYETBENINTKTEND 1423

QY 567 PDEGEIQRQHIVKINGKEVTPSPDPIQIKDQTKLNKIVWDPRART--TVK--EPILKDTG 622
 Db 1424 INEEVR-----NBEKRSINHIND-----TNINHIDBYNDTYPNFKDTECVHNENN 1472

QY 623 EYSELKPHRVVTIONKREMSSTIVSBEDEFLIPVYGELEKQYQFDGWEISPEGKHDAG 682
 Db 1473 MNSISQYTPVDTTRNHLVKR--NNQNP-----PEEGLNLBANFEEK--V 1516

QY 683 YYTINLSDD-----TPKPVFFCIEBKKERBENKPTP-----DVSCKCDN 720
 Db 1517 YIENNKKDHDHKGDSKYSSNLTSALNTICKSENDHNERKENTVVRKGEKGKIRKVKSGRN 1576
 QY 721 PQVN-HSOLN-----EFSRKEDLQREHRSOKS 746
 Db 1577 EKLNNEENYINNYDQMDNHRKNDITKKGENDBN 1609

Search completed: February 16, 2005, 19:19:52
 Job time : 188 secs

This Page Blank (uspto)

; LENGTH: 1787
 ; TYPE: PRT
 ; ORGANISM: K1 Parasite Clone
 US-10-4115-233-2

QY	740 BEHQKSDSLTKDVTAVLUDKNNISSKTTNPK 773
Db	:: : : : : : : : : : : : : 1761 TNKWEKT-----TKUDKNNKVKPQRRQQSK 1786

Search completed: February 16, 2005, 19:32:38
Job time : 147 secs

This Page Blank (uspto)

```

RESULT 15
US-09-949-016-1433
; Sequence 11433, Application US/09949016
; PATENT NO. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIORITY APPLICATION NUMBER: 60/241,755
; PRIORITY FILING DATE: 2000-10-20
; PRIORITY APPLICATION NUMBER: 60/237,768
; PRIORITY FILING DATE: 2000-10-03
; PRIORITY APPLICATION NUMBER: 60/231,498
; PRIORITY FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 201012
; SOFTWARE: PasteSeq for Windows Version 4.0
; SEQ ID NO 11433
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-1433

Query Match          4.6%;  Score 185
Best Local Similarity 20.0%;  Pred. No.
Matches 170;  Conservative 150;  Mismatch
Oy 21 LKQDGTGVERHHKHNRE--SIEKEKQSPITI
Db 1319 LKERIAGLLEBBKQKNEFQSQTLENKNTT
Oy 76 FTSETGKRMEEBYDQYDD-----
Db 1378 ELSRVTVKLUETAEKKDLEBRKQMLAKW
Oy 123 VLSPSKQDGHRFLIGKISKNVSKNATVY-----
Db 1438 CVSSLBEEBKQOLVKVKTQYSEBIRKYLEK
Oy 161 TRKD--FHSKTMFLYANINDVDSLAR
Db 1498 LOKDCDTCYQEKISALERTVKALE-----
Oy 217 KSEPVYVSYGNVIEGEGDLSKQRPDNJL
Db 1545 KQGQALBLASPKVUDDTOSBAAVLLDN
Oy 273 -GWAHKVTTYNPGKUDMGLSGNGVSKEDIA
Db 1605 BEKKHLBKKNMOKBDALRKVKHJBBTIG
Oy 312 QSVIMSAJGPNIRYQVETFKMDKGEAM
Db 1650 ---LQENLDS-TVYQLAFTPKMS---SIN
Oy 392 VEA1--KEIG-SMFLDTKPVNLSDKDNFV
Db 443 GGFHWELRUVNESVVDVNYLTYGDLHIDNTRD
Oy 1757 OGENTELL-----SQRSTR
Db 1697 EBBTRKLKDNCVSKDQLOMSINRTELK
Oy 503 DMD-----GNV--YIQTGTSIDL-NW
Db 1797 DLSNSLRLCKECKEOKANLBJGJTRQBDIIONSS
Oy 538 NIDPKGNTSIEYADGK--SVVFNENDKRN
Db 1850 BNNREOKLISLSSGKREBALQVALRQO

```

QY 593 IDKUUNIKUVKUANITVSKT LNKUTGS
Db 1909 VDKTNQMLBTTKTTKKEENTQKAOQDPSVKT
QY 651 DPTLWYKQBLKGTQFDEWBISPEG-----
Db 1969 DOLIOBAALENKQJB---BIRGLRSMDDD
QY 692 FIKPVFK-KIBKQREENKPTFDSKKCONP
Db 2024 QOKQILKVOQNBLENKXAKZBKLXSE
QY 746 ---SDSTKQVTA 754
Db 2083 VSISQUTQVTA 2094

Search completed: February 16, 2005, 19:20:41
Job time : 50 SECs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Om protein - protein search, using SW model

Run on: February 16, 2005, 19:06:53 ; Search time 170 Seconds

(without alignments)
1758.622 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGHTAESKPKNLGNGKEGS.....ATVLDKNNISSKSITNNPK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_GeneSeq 16Dec04;*

- 1: geneseqD19808;*
- 2: geneseqD19908;*
- 3: geneseqD20008;*
- 4: geneseqD20018;*
- 5: geneseqD20028;*
- 6: geneseqD20038;*
- 7: geneseqD2003bs;*
- 8: geneseqD20048;*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	4026	100.0	773	4 AAB8343
2	4026	100.0	2140	6 ABU01020
3	4026	100.0	2140	6 ABU5746
4	4026	100.0	2140	8 ADM02113
5	4015.5	99.7	2120	3 AAY81710
6	3783.5	94.1	2138	8 ADK48759
7	3048	75.7	637	8 ADR84534
8	750.5	18.6	1529	8 ADR6136
9	615	15.3	117	2 ABP55096
10	615	15.3	117	5 ABP54590
11	615	15.3	117	7 ADC51549
12	227	5.6	1196	6 ABU24813
13	222	5.5	861	7 ABO23608
14	220	5.5	1166	8 ADK9186
15	218	5.4	1639	2 AAW54145
16	218	5.4	1639	5 AAE29345
17	217	5.4	1166	8 ADK9185
18	217	5.4	1166	8 ADK9178
19	217	5.4	1166	8 ADK9184
20	216	5.4	1166	8 ADK9183
21	216	5.4	1233	5 ABP50203
22	216	5.4	1233	5 ABP29675
23	216	5.4	1233	8 ADK91765
24	216	5.4	1233	8 ADP10469
25	216	5.4	1239	5 ABP25822

ALIGNMENTS

RESULT 1	
ID	AAB48343
XX	AAB48343 standard; protein; 773 AA.
AC	AAB48343;
XX	
DT	20-APR-2001 (first entry)
XX	
DB	S. pneumoniae Sp130 polypeptide.
XX	
KW	Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal; KW bronchial; lung; blood; infection; immune response; immunotherapy; KW antibacterial; auditory; vaccine.
XX	
OS	S. pneumoniae.
XX	
PD	21-DEC-2000.
XX	
PF	09-JUN-2000; 2000WO-US015925.
XX	
PR	10-JUN-1999; 99US-0138433P.
XX	
(MEDI -) MED IMMUNE INC.	
PA	
XX	
PI	Adamou JE, Choi GH;
XX	
DR	WPI; 2001-112197/12.
DR	N-PSDB; AAC84742.
XX	
PS	New vaccines comprising Sp128 or Sp130 polypeptides, for treating and preventing pneumococcal infections, particularly infections caused by Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or blood infections.
PT	Claim 8; Page 51-54; 54pp; English.
XX	
CC	The invention relates to novel immunogenic polypeptides, Sp128 and Sp130 from S. pneumoniae. Vaccines comprising the polypeptides are useful for the treatment and prevention of pneumococcal infections, particularly infections caused by Streptococcus, such as otitis media, nasopharyngeal, bronchial, lung or blood infections. The antigens are used as immunogenic agents to stimulate an immune response. The antisera and antibodies may also be used in diagnosing and treating pneumococcal infections.
CC	Recombinant polypeptide serve as a mechanism for stimulating production of antibodies for use in passive immunotherapy, diagnostic reagents, and as reagents in other processes such as affinity chromatography. The

CC	present sequence represents the <i>S. pneumoniae</i> Sp130 polypeptide	KW	antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
XX		XX	gene therapy; vaccine.
SQ	Sequence 773 AA;	OS	Streptococcus pneumoniae; type 4 strain.
Query Match	100.0%; Score 4026; DB 4; Length 773;	XX	
Best Local Similarity	100.0%; Pred. No. 7.5e-25;	OS	
Matches	773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	XX	
QY	1 KLGELAESKPKNGLGNGKEGSLKKDTGVEHHQENNESEIKENSSFTIDRNISTRDENK	60	PN
DB	1 KLGELAESKPKNGLGNGKEGSLKKDTGVEHHQENNESEIKENSSFTIDRNISTRDENK	60	XX
QY	61 DIKKLIKKEPKREVDIFSETGKMEYDYKDDKGNIAIADDGDTLVEYETEKLDEIKSKI	60	PD
DB	61 DIKKLIKKEPKREVDIFSETGKMEYDYKDDKGNIAIADDGDTLVEYETEKLDEIKSKI	60	XX
QY	61 DIKKLIKKEPKREVDIFSETGKMEYDYKDDKGNIAIADDGDTLVEYETEKLDEIKSKI	60	PP
DB	61 DIKKLIKKEPKREVDIFSETGKMEYDYKDDKGNIAIADDGDTLVEYETEKLDEIKSKI	60	XX
QY	121 YGTLSPSKDGHFILGKISNVNSKNAVYQYNNKSTIEKATKVDPSKTMPTDLYANIND	180	PR
DB	121 YGVLSPSKDGHFELGKISNVNSKNAVYQYNNKSTIEKATKVDPSKTMPTDLYANIND	180	XX
QY	181 IVGLAFAGDMRLPVKNDQKKEIKURMPKEKETSEYPTVSYGNVIEGEGDLSK	240	XX
DB	181 IVGLAFAGDMRLPVKNDQKKEIKURMPKEKETSEYPTVSYGNVIEGEGDLSK	240	PA
QY	241 KPNLTKNESGKLYSDERQOYLKONITRAGYALKPTVNGKTMLENGVSYED	300	(CHIR-) INST GENOMIC RES.
DB	241 KPNLTKNESGKLYSDERQOYLKONITRAGYALKPTVNGKTMLENGVSYED	300	XX
QY	301 AKLQANPNLRASETTYADSRSNVEDGRSTOSVMSALDGENIIRYQVFTKNDGK	300	PI
DB	301 AKLQANPNLRASETTYADSRSNVEDGRSTOSVMSALDGENIIRYQVFTKNDGK	300	Masignani V, Tettelin H, Fraser C;
QY	361 IDKGNLVTDSSKVLKFGDKDKEYTGEDKFWAEGDSMFLIDTPEVNLMSMDK	420	XX
DB	361 IDKGNLVTDSSKVLKFGDKDKEYTGEDKFWAEGDSMFLIDTPEVNLMSMDK	420	DR
QY	420 421 SKSNKITYRNPFLYRKGKSDKGPMWLRVWSVWNLTYGDLHDNTDNPNIKAVK	480	WPI; 2003-040579/03.
DB	421 SKSNKITYRNPFLYRKGKSDKGPMWLRVWSVWNLTYGDLHDNTDNPNIKAVK	480	N-PSDB; ABX06302.
QY	421 SKSNKITYRNPFLYRKGKSDKGPMWLRVWSVWNLTYGDLHDNTDNPNIKAVK	480	XX
DB	421 SKSNKITYRNPFLYRKGKSDKGPMWLRVWSVWNLTYGDLHDNTDNPNIKAVK	480	PS
QY	481 DGDIMDWMKDYKANGFDDKUTMDGNYLQGQYSDANAKAVGVMHOFPLYVKPEVNTD	540	XX
DB	481 DGDIMDWMKDYKANGFDDKUTMDGNYLQGQYSDANAKAVGVMHOFPLYVKPEVNTD	540	CC
QY	540 541 PKGNTSTYEADGSKVVENINDKNGDFGEQOHTYINGKEYTSENDIKQIDTNIK	600	CC
DB	540 541 PKGNTSTYEADGSKVVENINDKNGDFGEQOHTYINGKEYTSENDIKQIDTNIK	600	CC
QY	600 601 IWKDPRANTTKEPLANKTGESELKPRHTVTLNGKMSSTVSEEDLILPKGE	660	CC
DB	600 601 IWKDPRANTTKEPLANKTGESELKPRHTVTLNGKMSSTVSEEDLILPKGE	660	CC
QY	660 661 LKGYQFDGWEISGFEGKDDAGYVNTSKDTPTKPVFRKKEERENKPTFDVSKKD	720	CC
DB	660 661 LKGYQFDGWEISGFEGKDDAGYVNTSKDTPTKPVFRKKEERENKPTFDVSKKD	720	CC
QY	720 721 PONHSQLNESEKLDQREIISQKSUTKDYATVLDKNISSKSTNNPK	773	CC
DB	720 721 PONHSQLNESEKLDQREIISQKSUTKDYATVLDKNISSKSTNNPK	773	CC
RESULT 2			
ABU01020	Query Match	100.0%; Score 4026; DB 6; Length 2140;	XX
ID	Best Local Similarity	100.0%; Pred. No. 2.9e-24;	OS
ABU01020;	Matches	773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	XX
XX	AC	AC ABU01020;	XX
XX	DT	23-OCT-2003 (revised)	XX
XX	DT	11-FEB-2003 (first entry)	XX
DE	S. pneumoniae type 4 strain protein from coding region #590.	DE	61 DIKKLIKKEPKREVDIFSETGKMEYDYKDDKGNIAIADDGDTLVEYETEKLDEIKSKI
XX	Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;	XX	120
KW		DB	1394 DIKKLIKKEPKREVDIFSETGKMEYDYKDDKGNIAIADDGDTLVEYETEKLDEIKSKI
			1453

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KJ, Zyskind JW;
 PT Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX DR WPI; 2003-02926/02.
 DR -PSDB; ACR49616.

PT New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 73670; 176pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that inhibits proliferation of an organism; (8) manufacturing an antibiotic; (9) profiling a compound or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (10) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp://wipo.int/pub/published_pct_sequences

XX SQ Sequence 2140 AA;

Query Match 100.0%; Score 4026; DB 6; Length 2140;
 Best Local Similarity 100.0%; Pred. No. 2.9e-224; Mismatches 0; Index 0; Gaps 0;

AC Matches 773; Conservative 0; MisMatches 0; Index 0; Gaps 0;

AC ABU4546;

XX DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #31273.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS *Streptococcus pneumoniae*.

XX FN WO2002771183-A2.

XX PD 03-OCT-2002.

XX PR 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948933.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

PA (BLIT-) BLITRA PHARM INC.

RESULT 3

ABU4546 standard; protein: 2140 AA.

ABU4546;

1 KUGETABSKFKLNGNGEGLSKKDTGVEHHHOENESTESIKERKSFTIDRNISTIDRFENK 60

1334 KUGEBIAKSFKFLNGNGEGLSKKDTGVEHHHOENESTESIKERKSFTIDRNISTIDRFENK 1393

61 DZKLLKKEFREDFSETGKMEEDYKDGGNIAVYDGTDLIYETEKLEDEKSKI 120

1394 DZKLLKKEFREDFSETGKMEEDYKDGGNIAVYDGTDLIYETEKLEDEKSKI 1453

121 YGULSPSKDGHFELGKISNSKNAKYGGNYKSEIKATKDFSKTMTDLYANIND 180

1514 YGULSPSKDGHFELGKISNSKNAKYGGNYKSEIKATKDFSKTMTDLYANIND 1513

241 KDPNLTKMESGKLYSDSEKQQLKONILRGYALKVTTNGKTMLEGNGVSKEDI 300

1574 KDPNLTKMESGKLYSDSEKQQLKONILRGYALKVTTNGKTMLEGNGVSKEDI 1633

301 AKIQKAMNRLASETTYIADSRNVEDGRSTOSVMSALDGENIIRYQVFTKMDRGEA 360

1634 AKIQKAMNRLASETTYIADSRNVEDGRSTOSVMSALDGENIIRYQVFTKMDRGEA 1693

xx SQ Sequence 2140 AA;

xx Best Local Similarity 100.0%; Score 4026; DB 8; Length 2140; Matches 773; Conservative 0; 0; Mismatches 0; Indels 0; Gaps 0;

Qy 361 IDKGDNLYTDSKLVLRGKDKEYTGEDKFNVAIKEDGSMFLPDTKPVNLSMDKMYFNP 420

Db 1694 IDKDGNLVTDSSKLVLFGKDKEYTGEDKFNVAIKEDGSMFLPDTKPVNLSMDKMYFNP 1753

Qy 421 SKSNKTYVRNPEFTYLRGKISDKGGFVNWLVRVNESWDNTLYLGDHLIDNTDRDNFLNKVK 480

Db 1754 SKSNKTYVRNPEFTYLRGKISDKGGFVNWLVRVNESWDNTLYLGDHLIDNTDRDNFLNKVK 1813

Qy 481 DGDIMDGWGMKDYKANGFPDKVTDMDGNYLQTGYSIDLNAKAVGAVHYPFLDNVKPEVNID 540

Db 1814 DGDIMDGWGMKDYKANGFPDKVTDMDGNYLQTGYSIDLNAKAVGAVHYPFLDNVKPEVNID 1873

Qy 541 PKGNTSIEYADGKSVVFINDKRNNGFDGEI0B0H1YINGKEYTSFNDIKOIDKTLN1K 600

Db 1874 PKGNTSIEYADGKSVVFINDKRNNGFDGEI0B0H1YINGKEYTSFNDIKOIDKTLN1K 1933

Qy 601 IWKDFAKNTTKEFINKDGTAVSEUKPRHTVQNGKEMSTIVSEEDFLPVYKG 660

Db 1934 IWKDFAKNTTKEFINKDGTSEVSEUKPRHTVQNGKEMSTIVSEEDFLPVYKG 1993

Qy 661 LEKGYQFDGWEISGPEGKKGADGVINLISKDFTIKPVKKEKBEENKPTFDVSKKD 720

Db 1994 LEKGYQFDGWEISGPEGKKGADGVINLISKDFTIKPVKKEKBEENKPTFDVSKKD 2053

Qy 721 PQVNHSQLNESHSRKEDLQREHSQKSDSTKDVATVLDKRNINSSKSTNNPK 773

Db 2054 PQVNHSQLNESHSRKEDLQREHSQKSDSTKDVATVLDKRNINSSKSTNNPK 2106

RESULT 4

ADM9213

ID ADM92113 standard; protein; 2140 AA.

XX AC ADM92113;

XX DT 03-JUN-2004 (first entry)

xx S pneumoniae antigenic protein sequence SeqID310.

xx KW antibacterial; gene therapy; *Streptococcus pneumoniae* infection; antigenic.

xx DT 11-MAR-2004.

xx WO2004020609-A2.

xx PP 02-SEP-2003; 2003WO-US027401.

xx PR 30-AUG-2002; 2002US-0407082P.

xx PA (TUFT) UNIV TUFTS.

xx PT Camilli A, Hava DL;

xx DR WPI: 2004-239189/22.

xx DR N-PSDB; ADM91876.

xx PT New *Streptococcus pneumoniae* nucleic acid molecules, useful for diagnosing, treating and preventing active infections of *Streptococcus pneumoniae*.

xx PS Claim 27; SEQ ID NO 310; 123pp; English.

xx CC This invention relates to novel isolated *Streptococcus pneumoniae* nucleic acid molecules and the antigenic polypeptides encoded by them. The invention may be useful for the production of compounds with an antibacterial activity or for gene therapy. The nucleic acid molecules, compositions and methods disclosed are useful for treating *Streptococcus pneumoniae* infection. The present sequence is that of an *S pneumoniae* protein of the invention.

xx

xx SQ Sequence 2140 AA;

xx Best Local Similarity 100.0%; Score 4026; DB 8; Length 2140; Matches 773; Conservative 0; 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KUGELIAESFKNGLNGKGSLSKDKTQGVEHHQENESIKEKSSFTIDRNISTIDRFENK 60

Db 1334 KUGELIAESFKNGLNGKGSLSKDKTQGVEHHQENESIKEKSSFTIDRNISTIDRFENK 1393

Qy 61 DIKKLIKKPKEVDDTSETGKRMETDYKDDKGNIAIDGTDIYEYKELDETKS1 120

Db 1334 DILKLUKKRKEVDDTSETGKRMETDYKDDKGNIAIDGTDIYEYKELDETKS1 1453

Qy 121 YGVLSRSPKDGHEIIGKISNSKAVYQYNNKSYEIKATKDFHSKTMFDLYANID 180

Db 1454 YGVLSRSPKDGHEIIGKISNSKAVYQYNNKSYEIKATKDFHSKTMFDLYANID 1513

Qy 181 IVDGLAFAGDKMLFLVKNDOKKAEIKRMPKEKETSEPYVSSGVNLBEGEDLSK 240

Db 1514 IVDGLAFAGDKMLFLVKNDOKKAEIKRMPKEKETSEPYVSSGVNLBEGEDLSK 1573

Qy 241 KPDNLTMESKSYISSESKBQOYLKONILRKGYALKVTTNGKIDMLEGNGVSKEDI 300

Db 1574 KPDNLTMESKSYISSESKBQOYLKONILRKGYALKVTTNGKIDMLEGNGVSKEDI 1633

Qy 301 AKIOKAPNRLASETTYADSRSNVEGRSTQSVIISALDGFIIRYQVFTRMDKGEA 360

Db 1634 AKIOKAPNRLASETTYADSRSNVEGRSTQSVIISALDGFIIRYQVFTRMDKGEA 1693

Qy 361 IDKGDNLYTDSKLVLRGKDKEYTGEDKFNVAIKEDGSMFLPDTKPVNLSMDKMYFNP 420

Db 1694 IDKDGNLVTDSSKLVLFGKDKEYTGEDKFNVAIKEDGSMFLPDTKPVNLSMDKMYFNP 1753

Qy 421 SKSNKTYVRNPEFTYLRGKISDKGGFVNWLVRVNESWDNTLYLGDHLIDNTDRDNFLNKVK 480

Db 1754 SKSNKTYVRNPEFTYLRGKISDKGGFVNWLVRVNESWDNTLYLGDHLIDNTDRDNFLNKVK 1813

Qy 481 DGDIMDGWGMKDYKANGFPDKVTDMDGNYLQTGYSIDLNAKAVGAVHYPFLDNVKPEVNID 540

Db 1814 DGDIMDGWGMKDYKANGFPDKVTDMDGNYLQTGYSIDLNAKAVGAVHYPFLDNVKPEVNID 1873

Qy 541 PKGNTSIEYADGKSVVFINDKRNNGFDGEI0B0H1YINGKEYTSFNDIKOIDKTLN1K 600

Db 1874 PKGNTSIEYADGKSVVFINDKRNNGFDGEI0B0H1YINGKEYTSFNDIKOIDKTLN1K 1933

Qy 601 IWKDFAKNTTKEFINKDGTAVSEUKPRHTVQNGKEMSTIVSEEDFLPVYKG 660

Db 1934 IWKDFAKNTTKEFINKDGTSEVSEUKPRHTVQNGKEMSTIVSEEDFLPVYKG 1993

Qy 661 LEKGYQFDGWEISGPEGKKGADGVINLISKDFTIKPVKKEKBEENKPTFDVSKKD 720

Db 1994 LEKGYQFDGWEISGPEGKKGADGVINLISKDFTIKPVKKEKBEENKPTFDVSKKD 2053

Qy 721 PQVNHSQLNESHSRKEDLQREHSQKSDSTKDVATVLDKRNINSSKSTNNPK 773

Db 2054 PQVNHSQLNESHSRKEDLQREHSQKSDSTKDVATVLDKRNINSSKSTNNPK 2106

RESULT 5

AY81710

ID AY81710 standard; protein; 2120 AA.

XX AC AY81710;

XX DT 02-JUN-2000 (first entry)

xx Streptococcus pneumoniae protein sequence ID3.

xx DR Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS; bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism; kidney disease; diabetes; immunosuppressive disorder; otitis media;

KW pneumococcal septicaemia; sinusitis; meningitis; therapy.
XX
OS Streptococcus pneumoniae.
XX
PN WO2006738-A2.
XX
PD 10-FEB-2000.
XX
PF 27-JUL-1999; 99WO-GB002452.
XX
PR 27-JUL-1998; 98GB-00016336.
PR 19-MAR-1999; 99US-0125339P.
XX
PA (MICR-) MICROBIAL TECHNICS LTD.
XX
PT Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
XX
DR WPI; 2000-195301/17.
XX
PT Streptococcal proteins and polynucleotides useful for diagnosis, treatment and prophylaxis of bacterial infections.
XX
PS Claim 2; Page 41-42; 76pp; English.

This sequence represents a Streptococcus pneumoniae protein of the invention. The proteins (or their homologues, derivatives and/or fragments) are useful as immunogens or antigens. Immunogenic or antigenic compositions comprising the proteins are useful as vaccines and also in diagnostic assays. The sequences are useful for the detection or diagnosis of S. pneumoniae infection, by contacting a sample to be tested with them. Agents capable of antagonising, inhibiting or interfering with the function or expression of the protein or polypeptide are useful in medical compositions in the treatment or prophylaxis of S. pneumoniae infection. As the sequences can be used to treat S. pneumoniae infection, they can be used to treat bacterial pneumonia, which has high rates in young children, the elderly, and in patients with predisposing conditions such as asplenia, heart, lung and kidney disease, diabetes, alcoholism, or with immunosuppressive disorders, especially AIDS. They can also be used to treat pneumococcal septicaemia, otitis media, sinusitis, and meningitis.

XX
SO Sequence 2120 AA;

Query	Match	Score	DB	Length	
Best Local Similarity	99.9%	4015.5	3	2120;	
Matches	773;	Conservative	0;	Mismatches	0;
				Indels	1;
				Gaps	1;

XX
QY 1 KLGEGIAASKFKFLNGLNGKEGSKLKDGTGVEHHRQENESIKEKSFTIDRNISTRDENK 60
DB 1313 KLGEGIAESKFKNLNGKEGSKLKDGTGVEHHRQENESIKEKSFTIDRNISTRDENK 1372
QY 61 DLUKLKIKKFREYDDPSETGKMEREDYKDDKGNTIAYDDGTLEYETEKLDEIKSKI 120
DB 1373 DLUKLKIKKFREYDDPSETGKMEREDYKDDKGNTIAYDDGTLEYETEKLDEIKSKI 1432
QY 121 YGVGLSPSKDGHFETLGKISNVSKNAKVYVGNNSKIEETKAKDIFHSKMTMDLYANIND 180
DB 1433 YGLSPSKDGHFETLGKISNVSKNAKVYVGNNSKIEETKAKDIFHSKMTMDLYANIND 1492
QY 181 YVGLLAGFAGDMRLFVKDQDKKAIEKIRMPKETKETSEPVYSSVGNVILEGGEDLSK 240
DB 1493 YVGLLAGFAGDMRLFVKDQDKKAIEKIRMPKETKETSEPVYSSVGNVILEGGEDLSK 1552
QY 241 KPNLTKQESGKYDSKQMLKDNTLTKYALKTTYKGTOMLEGVYSEDI 300
DB 1553 KPNLTKQESGKYDSKQMLKDNTLTKYALKTTYKGTOMLEGVYSEDI 1612
QY 301 AKIQKANPLRALSETIYIADSRAVEDGRSTOSVMSALDGENIIRQVFFKNDIGEA 360
DB 1613 AKIQKANPLRALSETIYADSRAVEDGRSTOSVMSALDGENIIRQVFFKNDIGEA 1672
QY 361 IDKQGNLVTDSSKVLVLFGKDKEYTGEDGSMPLIDTKPVNLSDMDKNYFNP 420

Db 1673 IDKQGNLVTDSSKVLVLFGKDKEYTGEDGSMPLIDTKPVNLSDMDKNYFNP 1732
QY 421 SGSNKTYVRNPFPYLRKISDKGGFWNLRLNESVUDNLYQGDLMIDNTR-DENIKLNV 479
Db 1733 SGSNKTYVRNPFPYLRKISDKGGFWNLRLNESVUDNLYQGDLMIDNTR-DENIKLNV 1792
Db 1792 KQDIDMMGKQYKANGPPDKUTDMGCVYLOTGKSDLNAKAVGVMQFLDNVKPEVNI 539
QY 480 KQDIDMMGKQYKANGPPDKUTDMGCVYLOTGKSDLNAKAVGVMQFLDNVKPEVNI 1852
Db 1793 KQDIDMMGKQYKANGPPDKUTDMGCVYLOTGKSDLNAKAVGVMQFLDNVKPEVNI 1852
QY 540 DPKGNTSTIEYADGKSVVFNTINDKRNNGPDLGELOEQIYIYNGKBYTFRNDIKOIIDKTLNT 599
Db 1853 DPKGNTSTIEYADGKSVVFNTINDKRNNGPDLGELOEQIYIYNGKBYTFRNDIKOIIDKTLNT 1912
Db 600 KLUVKDPAARNTTKEFLINKOTGEVSLKPRVVTIQLNGKMSSTIUSERDPLVYKG 659
QY 1913 KLUVKDPAARNTTKEFLINKOTGEVSLKPRVVTIQLNGKMSSTIUSERDPLVYKG 1972
Db 660 ELEKGYQPDGWAISSGFFGKKGQDGYVILNLSKTFIKPVFKKLEBKGEBENKTFDVSKKD 719
Db 1973 ELEKGYQPDGWAISSGFFGKKGQDGYVILNLSKTFIKPVFKKLEBKGEBENKTFDVSKKD 2032
QY 720 NPOVNHSQLNESHRKEDLQRBRHSOKSDSTKDVATVLDKNNISSKSTTNHNPK 773
Db 2033 NPOVNHSQLNESHRKEDLQRBRHSOKSDSTKDVATVLDKNNISSKSTTNHNPK 2086

RESULT 6
ID ADK48759
XX ADK48759 standard; protein; 2138 AA.
AC ADK48759;
XX
DT 20-MAY-2004 (first entry)
XX Streptococcus pneumoniae protein, Seq ID No 5274.
XX Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
OS Streptococcus pneumoniae.
XX
PN US6699703-B1.
XX
PD 02-MAR-2004.
XX
PF 26-MAY-2000; 2000US-00583110.
XX
PR 02-JUL-1997; 97US-0051533P.
PR 12-MAY-1998; 98US-0085131P.
PR 30-JUN-1998; 98US-00107433.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Housewright CE;
XX
DR WPI; 2004-212393/20.
XX
N-PSDB; ADK46098.
XX
PT New nucleic acid molecules and polypeptides useful for diagnosing, preventing and treating pathological conditions resulting from bacterial infection, e.g. Streptococcus pneumoniae infection, and in drug screening.
XX
PS Disclosure; SEQ ID NO 5274; 301pp; English.
XX
CC The invention relates to isolated Streptococcus pneumoniae nucleic acids and polypeptides. The nucleic acids and proteins are useful for diagnosing, preventing and treating pathological conditions resulting from bacterial infection, such as S. pneumoniae infection. These may also be used for drug screening procedures. The present sequence represents a Streptococcus pneumoniae polypeptide of the invention. Note: The sequence data for this patent did not appear in the printed specification but was

C 1 P 0

5. 26.

CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 2138 AA:

Query Match Best Local Similarity 94.1%; Score 3789.5; DB 8; Length 2138;
 Matches 733; Conservative 15; Mismatches 24; Indels 1; Gaps 1;
 XX
 QY 1 KLGEGIABSKPKNQNGKGGSGLICKDTCEVHHQBNERSIKKSPTIDRNISTIRDPENK 60
 DB 1333 KLGEGIAPSKFKOLKVNQKDSLKETAEVENNLVDMQSIKSLFNLNHTKISTIROFENK 1392
 QY 61 DLKLLKKRERREVDDETFSETGERMEEVYDVKDQGNTIAYQFTKDMLEGNGVSKED 60
 DB 1393 DLKLLKKRERREVDDETFSETGERMEEVYDVKDQGNTIAYQFTKDMLEGNGVSKED 1392
 QY 121 YCVLSPSKDGHFBLGKISNNSKNAKYQYGNWKSBIKATKDFISKTMFDLYANIND 180
 DB 1452 YCVLSPSKDGHFBLGKISNNSKNAKYQYGNWKSBIKATKDFISKTMFDLYANIND 1511
 QY 181 IVGLATAGDMLFLVKNDQKAEIKTRMPKETKETSEPYPVSSYGNVIEGEGDLSK 240
 DB 1512 IVGLATAGDMLFLVKNDQKAEIKTRMPKETKETSEPYPVSSYGNVIEGEGDLSK 1571
 QY 241 KEDNLTKNESKSKYPSKSEQKQTLKONILRKGYALKVTTNPGBTMLEGNGVSKED 300
 DB 1572 KEDNLTKNESKSKYPSKSEQKQTLKONILRKGYALKVTTNPGBTMLEGNGVSKED 1631
 QY 301 AKIQKAMPLRALSETTYAASRNVEDGRSTQSIVMSALDGFNTIYQVFTPKMDKGEA 360
 DB 1632 AKIQKAMPLRALSETTYAASRNVEDGRSTQSIVMSALDGFNTIYQVFTPKMDKGEA 1691
 QY 361 IDKDGNAUTDSSKVLVKGDKDKEYTGEKSWEAKEGDSMFLDTKPVNLSMDKQYFNP 420
 DB 1692 IDKDGNAUTDSSKVLVKGDKDKEYTGEKSWEAKEGDSMFLDTKPVNLSMDKQYFNP 1751
 QY 421 SISKNKTYVRNPRLRCKISDKGGFWNLRLRNESVWDNLYLGGDHLIDNTDPNKLNVK 480
 DB 1752 SKSNKTYVRNPRLRCKISDKGGFWNLRLRNESVWDNLYLGGDHLIDNTDPNKLNVK 1811
 QY 481 DEDIDMPGMKQKANGPDKVTDPMGAVYLTGYSINAKVGHYQFLYVNUKPNVID 540
 DB 1812 DEDIDMPGMKQKANGPDKVTDPMGAVYLTGYSINAKVGHYQFLYVNUKPNVID 1871
 QY 541 PKGNTSRYDGKSVVNNINDKRNNGDGEQOHYINGKTSNDIKUDKLNLK 600
 DB 1872 PKGNTSRYDGKSVVNNINDKRNNGDGEQOHYINGKTSNDIKUDKLNLK 1931
 QY 601 IVVKDFARNTVKEFILNKDGEVSELPKPRVYTONGKEMSTIVSEEDPILPVKGE 660
 DB 1932 IVVKDFARNTVKEFILNKDGEVSELPKPRVYTONGKEMSTIVSEEDPILPVKGE 1991
 QY 661 LKGKQDGMWLSGPECKKDKAGYVWNLSDPFIKPKFKKRKELENKPKPFDVSKDN 720
 DB 1992 LKGKQDGMWLSGPECKKDKAGYVWNLSDPFIKPKFKKRKELENKPKPFDVSKDN 2051
 QY 721 PQVNHSLQNEISKREDQREPSQSKDSTKQVATDKNMISSTNNPK 773
 DB 2052 PQVNHSLQNEISKREDQREPSQSKDSTKQVATDKNMISSTNNPK 2104
 RESULT 7
 ADR94534
 ID ADR94534 standard; protein; 637 AA.
 XX
 AC ADR94534;
 XX
 DT 16-DEC-2004 (first entry)
 XX
 DE Novel S. pneumoniae protein sequence, SEQ ID 3169.
 XX
 KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;

RW bacterial infection.
 XX Streptococcus pneumoniae.
 OS Streptococcus pneumoniae.
 PN US6800744-B1.
 XX 05-OCT-2004.
 XX 30-JUN-1998; 98US-00107433. 6
 XX 02-JUL-1997; 97US-0051553P.
 PR 12-MAY-1998; 98US-0085131P.
 XX (GENO-) GENOME THERAPEUTICS CORP.
 PT Doucette-Stamm LA, Bush D;
 XX DR N-PSDB; ADR91931.
 XX WPI; 2004-697205/68.
 PT DR; ADR91931.
 XX Disclosure: SEQ ID NO 3169; 151PP; English.
 CC The invention relates to an isolated nucleic acid comprising a sequence
 CC encoding a Streptococcus pneumoniae ADR91369polypeptide, or its
 CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,
 ADR9485, ADR9800, ADR9487, ADR9499, ADR8523, ADR95642, ADR95682,
 ADR96079) or any of the fully defined sequences appearing as ADR91705,
 ADR9186, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or
 ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
 CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
 CC hybridisable under high stringency conditions to the nucleotide sequences.
 CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.
 CC Also included are a recombinant expression vector comprising the isolated
 CC nucleic acid cited above operably linked to a transcription regulatory
 CC element, a cell comprising the recombinant expression vector and a probe
 CC comprising at least 20 consecutive nucleotides of the nucleotide
 CC sequences as cited above. The methods and compositions of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC pathological conditions resulting from bacterial infection by
 CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
 CC otitis media. The present sequence is one of the 2603 disclosed S.
 CC pneumoniae protein sequences. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from USPTO at
 seqdata.uspto.gov/sequence.html?docid=6800744B1.
 XX Sequence 637 AA;

Query Match Best Local Similarity 75.7%; Score 3048; DB 8; Length 637;
 Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 XX
 QY 185 LAFAGDMRLFVTKNDQKAEIKTRMPKETKETSEPYPVSSYGNVIEGEGDLSKPN 244
 DB 151A PAGDMRLFVTKNDQKAEIKTRMPKETKETSEPYPVSSYGNVIEGEGDLSKPN 74
 QY 245 LTKMEGKISKDSKSEQKQTLKONILRKGYALKVTTNPGBTMLEGNGVSKED 304
 DB 75 LTKMEGKISKDSKSEQKQTLKONILRKGYALKVTTNPGBTMLEGNGVSKED 134
 QY 305 KANPNTRALSETTYAASRNVEDGRSTQSIVMSALDGFNTIYQVFTPKMDKGEA 364
 DB 135 KANPNTRALSETTYAASRNVEDGRSTQSIVMSALDGFNTIYQVFTPKMDKGEA 194
 QY 365 GNLVTDSSKVLVFLGKDKDKEYTGEKSWEAKEGDSMFLDTKPVNLSMDKQYFNP 424
 DB 195 GNLVTDSSKVLVFLGKDKDKEYTGEKSWEAKEGDSMFLDTKPVNLSMDKQYFNP 254
 QY 425 KIVRNPEFVLRGKISDKGGFWNLRLRNESVWDNLYLGGDHLIDNTDPNKLNVKGD 484

CC can be useful in vaccines for inducing protective antibodies against
 CC *Streptococcus pneumoniae*, for treatment or prevention of infection e.g.
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
 CC are used to detect *Streptococcus* infection (by usual hybridisation or
 CC amplification methods), also for isolating *Streptococcus* genes or their
 CC allelic variants. The protein can be used similarly to detect specific
 CC monitoring infections. Antibodies which bind the protein are used to
 CC detect corresponding antigens, to purify the protein and for passive
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000
 CC (especially 10-300) µg/ml per dose
 XX SQ Sequence 117 AA;

Query Match 15.3%; Score 615; DB 2; Length 117;
 Best Local Similarity 100.0%; Pred. No. 3.2e-28;
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQDGWEISGFGKDKDAGYVINVLSKDFIKPVKKIEKEENKPTFDVSK 716
 Db 1 YKGELEKGYQDGWEISGFGKDKDAGYVINVLSKDFIKPVKKIEKEENKPTFDVSK 60

QY 717 KKDNPQVNHSQLNESHRKEDLQREHQSKSISTKDVTATVLDKNNSSKSTNNPK 773
 Db 61 KKDNPQVNHSQLNESHRKEDLQREHQSKSISTKDVTATVLDKNNSSKSTNNPK 117

RESULT 10
 ID ABP54590 standard; protein; 117 AA.
 XX ABP54590;
 AC
 DT 04-SEP-2002 (first entry)
 XX S. pneumoniae SP043 protein sequence SEQ ID NO:68.
 DE Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
 KW antibacterial; Streptococcal infection; detection.
 OS Streptococcus pneumoniae.
 XX US2002061545-A1.
 XX
 PR 23-MAY-2002.
 XX
 PR 22-JAN-2001; 2001US-00765272.
 PR 30-OCT-1997; 97US-00961083.
 XX
 PA (CHOI) / CHOI G H.
 PA (KUNS) / KUNSCH C A.
 PA (BARA) / BARASH S C.
 PA (DILL) / DILLON P J.
 PA (DOUG) / DOUGHERTY B.
 PA (FANN) / FANNON M R.
 PA (ROSE) / ROSEN C A.
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;
 XX WPI; 2003-764547/72.
 DR N-PSDB; ADC45148.
 XX
 PT Novel polynucleotide encoding *Streptococcus pneumoniae* polypeptides
 PT useful for producing vaccines for prevention or attenuation of infection
 PT by *Streptococcus pneumoniae*.
 XX
 PS Example 1; SEQ ID NO 68; 58pp; English.
 XX
 CC The invention relates to an isolated polynucleotide consisting of a
 CC *Streptococcus pneumoniae* nucleic acid (appearing as ADC45122 and encoding
 CC SP028) one of 113 disclosed nucleic acids encoding 113 *S. pneumoniae*
 CC antigens. Also included are making a recombinant vector by inserting the
 CC nucleic acid into a vector, an isolated polynucleotide consisting of at
 CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
 CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
 CC acids are useful as DNA vaccine against *Streptococcus pneumoniae*
 CC infection (e.g. pneumonia). Nucleic acids derived from the *S. pneumoniae*
 CC antigen nucleic acids are useful as probes for use in diagnostic methods
 CC for detecting *S. pneumoniae* gene expression. The present sequence
 CC represents an *S. pneumoniae* antigenic protein.

CC *pneumoniae* antigens have antibacterial activity and can be used in
 CC vaccines. The *S. pneumoniae* antigens can also be used to prevent or
 CC attenuate a *Streptococcal* infection in an animal. The polynucleotides
 CC encoding the *S. pneumoniae* antigens can be used to detect *Streptococcus*
 CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
 CC of *S. pneumoniae* ORFs (open reading frames) which are used in an example
 CC from the present invention.

Query Match 15.3%; Score 615; DB 5; Length 117;
 Best Local Similarity 100.0%; Pred. No. 3.2e-28;
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQDGWEISGFGKDKDAGYVINVLSKDFIKPVKKIEKEENKPTFDVSK 716
 Db 1 YKGELEKGYQDGWEISGFGKDKDAGYVINVLSKDFIKPVKKIEKEENKPTFDVSK 60

QY 717 KKDNPQVNHSQLNESHRKEDLQREHQSKSISTKDVTATVLDKNNSSKSTNNPK 773
 Db 61 KKDNPQVNHSQLNESHRKEDLQREHQSKSISTKDVTATVLDKNNSSKSTNNPK 117

RESULT 11
 ID ADC45149
 ID ADC45149 standard; protein; 117 AA.
 XX
 AC ADC45149;
 XX
 DT 18-DEC-2003 (first entry)
 XX S. pneumoniae antigenic protein SP043.
 DR
 XX Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
 KW
 OS Streptococcus pneumoniae.
 XX
 PN US6573082-B1.
 PR 03-JUN-2003.
 XX
 PP 28-MAR-2000; 2000US-00536784.
 XX
 PN 31-OCT-1996; 96US-0029960P.
 PR 30-OCT-1997; 97US-00961083.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PT Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PT Rosen CA;
 XX WPI; 2003-764547/72.
 DR N-PSDB; ADC45148.
 XX
 PT Novel polynucleotide encoding *Streptococcus pneumoniae* polypeptides
 PT useful for producing vaccines for prevention or attenuation of infection
 PT by *Streptococcus pneumoniae*.
 XX
 PS Example 1; SEQ ID NO 68; 58pp; English.

CC The invention relates to an isolated polynucleotide consisting of a
 CC *Streptococcus pneumoniae* nucleic acid (appearing as ADC45122 and encoding
 CC SP028) one of 113 disclosed nucleic acids encoding 113 *S. pneumoniae*
 CC antigens. Also included are making a recombinant vector by inserting the
 CC nucleic acid into a vector, an isolated polynucleotide consisting of at
 CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
 CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
 CC acids are useful as DNA vaccine against *Streptococcus pneumoniae*
 CC infection (e.g. pneumonia). Nucleic acids derived from the *S. pneumoniae*
 CC antigen nucleic acids are useful as probes for use in diagnostic methods
 CC for detecting *S. pneumoniae* gene expression. The present sequence
 CC represents an *S. pneumoniae* antigenic protein.

ABQ8492 to ABQ84904 represents nucleic acids which encode the
 CC *Streptococcus pneumoniae* antigens given in ABP54557 to ABP54669. The S.

SQ Sequence 117 AA;

ABU24913 ID ABU24913 standard; protein; 1196 AA.

AC ABU24913;

XX DT 19-JUN-2003 (first entry)

DB Protein encoded by prokaryotic essential gene #1040.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Clostridium botulinum.

XX PN W0200277183-A2.

PD 03-OCT-2002.

XX PP 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-0094893.

PR 25-OCT-2001; 2001US-0342933P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zepkind JW, Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI: 2003-029926/02.

DR N-PSDB; AC22863.

XX PT New antisense nucleic acid, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 52737; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences)

SQ Sequence 1196 AA;

Query Match 5.6%; Score 227; DB 6; Length 1196;

Best Local Similarity 20.2%; Pred. No. 0.0021; Mismatches 185; Conservative 145; Indels 328; Gaps 46;

Matches 185; Conservative 145; Mismatches 260; Indels 328; Gaps 46;

DB 123 RLKDQG-LFMDTGIGKGYSITGQCKIAVLSGKPEERSLLEAGIVFKWKRKEAD 181

QY 1 KLGEBIABSKFKLNGNGEG-.....-SILKQDGTGVEHHHQENESI 39

Db 40 KEKSFIDRMRISTDPEN-KDQKLGKQKFREDF--TSEFGKRM-.....- 85

QY 182 KKLNS-T-TEONLIRIKQDILNTYERMPKLESEKAKFKLNUSEELRKKEVNMVYISID 239

Db 86 -BYDYKDDKGNIAYDDGTD-LEYETEKLBDIKSKYIGVLSPLSPSKDQHPEILGKISNSV 143

Db 240 IEKDLK-NISSNMLSGENTINLNKNEQSQYKEIISPK-.....-NEKLELDK--NNR 288

QY 144 NAKVYCCGNNYKSIETKATKTYDFHSKTMFDLYANINDIVDGLAFAGDMRLFVKNDQKCA 203

Db 289 NKEBYYNNKDNKDEI-.....-NENALIJKEKIKNLKD-.....-NIKRENTLN 331

Db 204 EIKR-.....-MPPKIKETKSEYV-.....- 223

Db 332 EBRLOQYKKEKEALEKKINKLKEEESYRERIGKKEDNVNPNKELKEEKLKLKILSEE 391

QY 224 --SSYGVIEGEGDL--SKQPDNLJTKMESGGKYS-.....-DSEKQOQLK 265

Db 392 IELSNVSNLNEISVNENETVNLENKLNDKNTKNSCISYISSINININTKEDIEKEKNIK 451

QY 266 DNIIIL-.....-RKQYALKVTTYNGFKTMDLEGNGVYSKD-.....-TAKIOK---AN 307

Db 452 ENILLKLNKNSRNISLKLISLNNEK-KLKEKPAYSRLEANTHMLSNLBKHVEGYN 510

QY 308 PNLRALESETTYIADSRSNVEDERSTOSVLSALDGFNIIYKVFTFONDKGEAIDKGNL 367

Db 511 RSVKTMHEVHSYGVKDNKGG-.....-CEVIGDITKVKKELETAMBIAGAI--SNV 560

QY 368 VTDSSKLVLFGKDKKEYTEDKPNVLAKEEDGMSMLFIDTKVNLSDMDKNYFNP--SRSN 424

Db 561 ITEDE-.....-NKKAKLNLWKKK-.....-SUGRATEPLPTIIGR 594

QY 425 KIVRNNEFYLRLGKISDKGGNWLRLRNVESVVDNYLYTGDIAIDNTRDFN1--KLNWDG 482

Db 595 KAKINN-.....-VTREDGP--LGIASLDLIDY-.....-DKFSNIDIVLGLTVALKD- 637

QY 483 DMDWGMKDYKANGFPDKVTDGNTYLTQTCYSDLAKAKAVGHYQFLYDNTKPEVNIDPK 542

Db 638 -MDSALKIARKLNUYSFKVITLEGEV-.....- 665

QY 543 GNT-.....-SIEYADGKSVFVNINDKRNFGDFGIQEHIYNGKEYTSFNDIKOILMLN 599

Db 666 GSLTGASLSKHRGSSI-.....-LKRK-.....-EIEE-.....-TKKELETKN- 700

QY 600 KIVVKDOPARTVYKEFLINKOTGEVSELPKPHRVVTIQLONGKEMSTIVSERDPILWYKG 659

Db 701 -.....-TIEEFGN-.....-ILLENKOKKI-.....-LDEENLNT-.....-KD 728

QY 660 ELEKGYQFDGWBISSGFFGKDA-.....-GYVNLSQLD-.....-TPIKPVFKKIEEKREENK 709

the native pattern of folding. Larger amounts of the protein can be

produced recombinantly than would be possible using the parasites as a source. (Updated on 25-MAR-2003 to correct PR field.)

Q Sequence 1639 AA;

Query Match Similarity 5.4%; Score 218; DB 2; Length 1639;
 Best Local Similarity 18.6%; Pred. No. 0.0011; Matched 140. Index 228. Game 19. March 157. Conservative 164. Mainstream 140.

Y 23 KDTTG-VEHHQOBENESTK-----EKSSTFDNRNSTTDFENDKQKJTKKCFREVDD 75
b 232 KDNVKGMDVYKKNKKTIENINELLEBESKKTIDKNGATKEE-----KKLYQAQY 283
Y 76 FTSETCKRMEEYDVKDKGNTIAYDGTOLYEYETKLDETKSKYLGWSPSKDGHFEL 135
b 284 DLSIYNGQLEE-----AHLISV-----LE-----KRIDTLK-----KNENKELL 319
b 136 GKSINVSXKAQVYGINNYSIEKATKTYDF-----SKWTFDLYANINDIVDGLA 188
b 320 DKINEKKNPPANSSTPNTLDDKNEKEIEKEIKAQKJTKENISLDFPL----- 373
b 189 GDMRLKTVKDNDOKKASIEKIRMPKETKTKSEVP-----YVSYGNV---TBGLG-GDLSK 239
b 374 -ELEVYLRKKN-KNIDISAKVETKESTEPEPNVYPNGVTPYLSYNDINNALNBLSFGDLI- 430
b 240 NKPDNUTKMESEGGKYSSEKQYL-LKDNTL-----RKGIAKVTYNPRTKDM 289
b 431 -NPFDYTKPEPSKNIYTDNERKCKFNEKEKCKKIESDKKSYEDRSKSIN-----DIT 485
b 290 EG-----NGVYSKE-----DIAKIQA-----NPNURALSTTIVADSRVNGDGRSTQV 334
b 486 KEYEKULNNEITDSKFNNTIDNTFEGOMGKRYSYKVEKLHTNTYASYENSKHNLAKLT 545
b 335 LMSALDGF-----NII-----RYQVFTEPKMDKGAEADKGDNLVTDSKLVFGKODKEYT 385
b 546 ALKYMEDYSLRNIVWEKELKVKYKNUJSKIELETFE- NIKKOBQLF-----EKIT 598
b 386 GEDKVNTEAILEDGSMFLFIDTRKPVNLSMDKNTYFNPSKSNSKLYVRFPEPFYLRGKTSDDKG 445
b 599 KDENKPDPEKILVEVDIIVKVQVQVQLL-----MNKIDELKKTOLILKVNVE-LKHNIHUPNSY 653
b 446 NWELRNTESVVDNYLIVGDLHIDNTRPNTKL-----NVK----- 480
b 654 KOENKOE-----PYYLULLKKESEIDKLUVKPMVKPEVSLINBEKENIKTEQGQSONSERPTEGEI 709
b 481 -----DGGDIMWGMKDYK----- 501
b 710 TQQATTKPGQGQAGSALEGDSTQAQAOEQKQAOQPPVPPVPEAKAQVTPPAPVNKTENV 769
b 502 TMDGMVYLYQTYGSDINA-----KAVGVMYQPLDVKPKEVNDPKGNTSLEYADGKSV 556
b 770 SKLD-----YLEKLYEFINTSYICHYKLYLVSHSTMNKEKLQKXITKEEKSLSQCDPLDL 826
b 557 FNIN-----DGRNNGFD-----EIOBQHUYINGKEYTSFNDIKQIDKTLNIKIV 603
b FNIQNNTPVMWSMFDLSLANSISOLFMEVYCEGMVCMNLYKLKDNDKLNBLBEEAKVYSTV 886
b 604 KDPARTNTVKERPLINKDTGEYSELKPHRTVTLQNGKEMSTIVSBDPFLPVYGELEK 663
b 887 KTLSSSSMOPSLTPODKPEVSANDTTSHTSTNLNSLKLLENFLS-----LGKNN 937
b 664 GYQFDGWEISGREGK-----KDAGYVVINLSKDTFIPVFKLCEBKGSEENKPTF-D 713
b 938 IYQ-----ELIGOKSSENIFYEKILKOSDFTFYNESFTNEVSKADDINSLDESKRKKLEED 993
b 714 VSKKKDPQVHNSQLMESRSHKEKDLOPESBHSQSKSDSTKDVTATVLDKNNISK-STNNP 771
b 994 INKLKXLTQLSFDLYNKKLKEREDKKKTGKQXQIKGTLKJKEQLSKLSKLNINP 1052

search completed: February 16, 2005, 19:16:44
ob time : 177 secs

Copyright (c) 1993 - 2005 Compugen Ltd.

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:10:40 ; Search time 45 Seconds

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues.

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/1/iaa/5A_COMB_pep:*

2: /cgn2_6/prodata/1/iaa/5B_COMB_pep:*

3: /cgn2_6/prodata/1/iaa/6A_COMB_pep:*

4: /cgn2_6/prodata/1/iaa/6B_COMB_pep:*

5: /cgn2_6/prodata/1/iaa/PCMS_COMB_pep:*

6: /cgn2_6/prodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3789.5	94.1	2138	4	US-09-583-110-5274	Sequence 5274, Ap
2	3048	75.7	637	4	US-09-107-433-3169	Sequence 3169, Ap
3	750.5	18.6	1529	4	US-09-107-433-4771	Sequence 4771, Ap
4	615	15.3	117	3	US-08-981-083-68	Sequence 68, Ap
5	615	11.7	117	4	US-09-536-384-68	Sequence 68, Ap
6	203	5.0	1166	4	US-09-200-650B-7	Sequence 7, Ap
7	200.5	5.0	1786	3	US-08-973-462-8	Sequence 8, Ap
8	191.5	4.8	2304	4	US-09-328-312-5821	Sequence 5821, Ap
9	188.5	4.7	670	4	US-09-107-433-4976	Sequence 4976, Ap
10	188.5	4.7	708	4	US-09-583-110-3019	Sequence 3019, Ap
11	187.5	4.7	1529	2	US-08-728-470-10	Sequence 10, Ap
12	187.5	4.7	1529	3	US-08-719-641-10	Sequence 10, Ap
13	186.5	4.6	930	4	US-09-200-650B-3	Sequence 3, Ap
14	186	4.6	1315	4	US-09-200-650B-5	Sequence 5, Ap
15	185	4.6	2733	4	US-09-949-016-1143	Sequence 11433, A
16	4.6	2529	5	US-09-949-016-6507	Sequence 6507, Ap	
17	184	4.6	1588	5	PCT-US93-07281-11	Sequence 11, Ap
18	184	4.6	1663	4	PCT-US93-0761-16	Sequence 16, Ap
19	181.5	4.5	778	3	US-09-134-001C-3868	Sequence 3868, Ap
20	179.5	4.5	746	4	US-09-710-277-652	Sequence 652, Ap
21	179	4.5	1183	2	US-08-447-031A-2	Sequence 2, Ap
22	179	4.4	2412	4	US-09-540-03459	Sequence 3459, Ap
23	178	4.4	2375	4	US-09-538-099-1131	Sequence 1131, Ap
24	177.5	4.4	1600	2	US-08-617-697-10	Sequence 10, Ap
25	177	4.4	800	6	5183745-3	Patent No. 5183745
26	177	4.4	800	6	5183745-3	Sequence 49, Ap
27	174.5					

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KUGRIBAESKFKNLGNKGEGS.....ATVLDKRNISSKSTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues.

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/1/iaa/5A_COMB_pep:*

2: /cgn2_6/prodata/1/iaa/5B_COMB_pep:*

3: /cgn2_6/prodata/1/iaa/6A_COMB_pep:*

4: /cgn2_6/prodata/1/iaa/6B_COMB_pep:*

5: /cgn2_6/prodata/1/iaa/PCMS_COMB_pep:*

6: /cgn2_6/prodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3789.5	94.1	2138	4	US-09-583-110-5274	Sequence 5274, Ap
2	3048	75.7	637	4	US-09-107-433-3169	Sequence 3169, Ap
3	750.5	18.6	1529	4	US-09-107-433-4771	Sequence 4771, Ap
4	615	15.3	117	3	US-08-981-083-68	Sequence 68, Ap
5	615	11.7	117	4	US-09-536-384-68	Sequence 68, Ap
6	203	5.0	1166	4	US-09-200-650B-7	Sequence 7, Ap
7	200.5	5.0	1786	3	US-08-973-462-8	Sequence 8, Ap
8	191.5	4.8	2304	4	US-09-328-312-5821	Sequence 5821, Ap
9	188.5	4.7	670	4	US-09-107-433-4976	Sequence 4976, Ap
10	188.5	4.7	708	4	US-09-583-110-3019	Sequence 3019, Ap
11	187.5	4.7	1529	2	US-08-728-470-10	Sequence 10, Ap
12	187.5	4.7	1529	3	US-08-719-641-10	Sequence 10, Ap
13	186.5	4.6	930	4	US-09-200-650B-3	Sequence 3, Ap
14	186	4.6	1315	4	US-09-200-650B-5	Sequence 5, Ap
15	185	4.6	2733	4	US-09-949-016-1143	Sequence 11433, A
16	4.6	2529	5	US-09-949-016-6507	Sequence 6507, Ap	
17	184	4.6	1588	5	PCT-US93-07281-11	Sequence 11, Ap
18	184	4.6	1663	4	PCT-US93-0761-16	Sequence 16, Ap
19	181.5	4.5	778	3	US-09-134-001C-3868	Sequence 3868, Ap
20	179.5	4.5	746	4	US-09-710-277-652	Sequence 652, Ap
21	179	4.5	1183	2	US-08-447-031A-2	Sequence 2, Ap
22	179	4.4	2412	4	US-09-540-03459	Sequence 3459, Ap
23	178	4.4	2375	4	US-09-538-099-1131	Sequence 1131, Ap
24	177.5	4.4	1600	2	US-08-617-697-10	Sequence 10, Ap
25	177	4.4	800	6	5183745-3	Patent No. 5183745
26	177	4.4	800	6	5183745-3	Sequence 49, Ap
27	174.5					

Title: US-09-583-110-5274

Perfect score: 4026

Sequence: 1 KUGRIBAESKFKNLGNKGEGS.....ATVLDKRNISSKSTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues.

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/1/iaa/5A_COMB_pep:*

2: /cgn2_6/prodata/1/iaa/5B_COMB_pep:*

3: /cgn2_6/prodata/1/iaa/6A_COMB_pep:*

4: /cgn2_6/prodata/1/iaa/6B_COMB_pep:*

5: /cgn2_6/prodata/1/iaa/PCMS_COMB_pep:*

6: /cgn2_6/prodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

RESULT 1

US-09-583-110-5274

; Sequence 5274, Application US/09583110

; Patent No. 6659703

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al.

; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics

; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics

; FILE REFERENCE: PATH00-07A

; CURRENT APPLICATION NUMBER: US/09/583, 110

; CURRENT FILING DATE: 2000-05-25

; PRIORITY APPLICATION NUMBER: US 09/107, 433

; PRIORITY FILING DATE: 1998-05-30

; PRIORITY FILING DATE: 1998-05-12

; PRIORITY APPLICATION NUMBER: US 60/051, 553

; NUMBER OF SEQ ID NOS: 5322

; SEQ ID NO 5274

; LENGTH: 2138

; TYPE: PRT

; ORGANISM: Streptococcus pneumoniae

US-09-583-110-5274

Query Match 94.1%; Score 3789.5; DB 4; Length 2138;

Best Local Similarity 94.8%; Pred. No. 1.1e-238; Mismatches 24; Indels 1; Gaps 1;

Matches 733; Conservative 15; MisMatches 24; Indels 1; Gaps 1;

Qy 1 KUGRIBAESKFKNLGNKGEGS.....ATVLDKRNISSKSTNNPNK 60

Db 1333 KUGRIBAESKFKNLGNKGEGS.....ATVLDKRNISSKSTNNPNK 1392

Qy 61 DIKKLIKKEFREDDTSSETGKMERDYKDDKGNIAVDDGTLIEYTKLDEISKI 120

Db 1393 DIKKLIKKEFREDDTSSETGKMERDYKDDKGNIAVDDGTLIEYTKLDEISKI 1451

Qy 121 YGVLSRSPSKDGHPEILGKISVSKNPKAYVYGNKSYSTEIKATVAKYDFHFSKTMWFDLYANIND 180

Db 1452 YGVLSRSPSKDGHPEILGKISVSKNPKAYVYGNKSYSTEIKATVAKYDFHFSKTMWFDLYANIND 1511

Qy 181 YVGLGLPAGDRLPKVNDKDKAEIKRMPKIKETSKYEPVYSSGNGVIELEGPSLKN 240

Db 1512 YVGLGLPAGDRLPKVNDKDKAEIKRMPKIKETSKYEPVYSSGNGVIELEGPSLKN 1571

Qy 241 KEDNLTKMESPCKYISDSEKQYLKONILRKYALKVTTNPKGKUDMLECGNGVSKED 300

Db 1572 KEDNLTKMESPCKYISDSEKQYLKONILRKYALKVTTNPKGKUDMLECGNGVSKED 1631

Qy 301 AKIQKANPRLASETTYIADSRSNEDGRSTQVLSALDFGPNLIRYQVTFKMDKGEA 360

C\8

Db 1632 AKIQKANPRLALSETTYADSRNVEDGRSTQAVIMSLADCFNIRYQVFPMNDKGEA 1691
 Qy 361 IDKDGNIUTDSKUVLFGKDKDKEYTGEDKFNVEAIKEDGSMFLPDTKPVNLSMDKRYFNP 420
 Db 1692 IDKDGNIUTDSKUVLFGKDKDKEYTGEDKSNVEAIKEDGSMFLPDTKPVNLSMDKRYFNP 1751
 Qy 421 SGSNKTYVRNPFYLGKISDKGGFNLWNRVNESVVDNYLYGDLAIDNTRDFNPKLNVK 480
 Db 1752 SKSNKTYVRNPFYLGKISDKGGFNLWNRVNESVVDNYLYGDLAIDNTRDFNPKLNVK 1811
 Qy 481 DGDIMDWMGKOKXKANGRPDKVTDMDGNYLQGYSQNLNAKAVGVHQLYDNVKPBNID 540
 Db 1812 DGDIMDWMGKOKXKANGRPDKVTDMDGNYLQGYSQNLNAKAVGVHQLYDNVKPBNID 1871
 Qy 541 PKGNTSIEYAGKSVFVNINDKRNNGPGEQOHIYINGKEYTSFNDIKOIDTKLNK 600
 Db 1872 PKGNTSIEYAGKSVFVNINDKRNNGPGEQOHIYINGKEYTSFNDIKOIDTKLNK 1931
 Qy 601 IWKDFAKNTTKEFLNKDGEVSELKPHRTVTIONKEMSTIVSEEDFILPVYKE 660
 Db 1932 IWKDFAKNTTKEFLNKDGEVSELKPHRTVTIONKEMSTIVSEEDFILPVYKE 1991
 Qy 661 YPKGYOPDGWELISGPGKDKDGYVNLSKDPTIKPVPKKLEKKEBNKPTFDVSKKDN 720
 Db 1992 YPKGYOPDGWELISGPGKDKDGYVNLSKDPTIKPVPKKLEKKEBNKPTFDVSKKDN 2051
 Qy 721 PQVNHSQLNESHRKELOREHSQKSDSTKDVATVLDKNNISSTNNPK 773
 Db 2052 PQVNHSQLNESHRKELOREHSQKSDSTKDVATVLDKNNISSTNNPK 2104

RESULT 2
 US-05-107-433-3169
 ; Sequence 3169, Application US/09107433
 ; Patent No. 6800744
 ; GENERAL INFORMATION:
 APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEAR ACID AND AMINO ACID
 SEQUENCES RELATING TO STREPTOCOCUS PNEUMONIAE FOR DIAGNOSTIC
 THERAPEUTICS

NUMBER OF SEQUENCES: 5206

CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 CITY: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02354

COMPUTER READABLE FORM:
 MEDIUM TYPE: CD/ROM ISO9660
 COMPUTER: <Unknown>
 OPERATING SYSTEM: <Unknown>
 SOFTWARE: <Unknown>

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/107-433
 FILING DATE: 30-Jun-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/ 095131
 FILING DATE: May 12, 1998
 APPLICATION NUMBER: 60/051553
 FILING DATE: July 2, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Arinello, Pamela Deneke
 REGISTRATION NUMBER: 40,489
 REFERENCE DOCKET NUMBER: GTC-011
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (781)893-5007
 TELEFAX: (781)893-8277
 INFORMATION FOR SEQ ID NO: 3169:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 637 amino acids
 TYPE: amino acid
 TOPOLOGY: linear

RESULT 3
 US-05-107-433-4771
 ; Sequence 4771, Application US/09107433
 ; Patent No. 6800744
 ; GENERAL INFORMATION:
 APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEAR ACID AND AMINO ACID
 SEQUENCES RELATING TO STREPTOCOCUS PNEUMONIAE FOR DIAGNOSTIC
 THERAPEUTICS

NUMBER OF SEQUENCES: 5206

CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 CITY: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02354

COMPUTER READABLE FORM:

MOLECULE TYPE: protein.
 HYPOTHETICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: Streptococcus pneumoniae
 FEATURE:
 NAME/KEY: misc. feature
 LOCATION: (B) LOCATION 1.. 637
 SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
 ; US-09-107-433-3169
 ; Query Match 75.7%; Score 3048; DB 4; Length 637;
 Best Local Similarity 99.3%; Pred. No. 5.6e-19; 1; Indels 0; Gaps 0;
 Matches 585; Conservative 3; Mismatches 1;
 Query 185 LAPAGDMRLFKNDKOKAETKIRKPEKIKETSKSYPYQSYGNVIELEGSDISKPKDN 244
 Db 75 LTKWESGKISITDSEKQYQJLKDNTIRKGALKVITYNPQKTDMLLEGNGYTSKEDIKIQ 134
 Query 305 KANPNLRLASBTTYADSRNVEDGRSTQAVIMSLADGFNTYQVFPMNDKRYFNPFSK 424
 Db 135 KANPNLRLASBTTYADSRNVEDGRSTQAVIMSLADGFNTYQVFPMNDKRYFNPFSK 194
 Query 365 GNLVTDSSKUVLFGKDKDKEYTGEDKSNVEAIKEDGSMFLPDTKPVNLSMDKRYFNPFSK 424
 Db 195 GNLVTDSSKUVLFGKDKDKEYTGEDKSNVEAIKEDGSMFLPDTKPVNLSMDKRYFNPFSK 254
 Query 425 KIYVRNPEFYLGKISDKGGFNLWNRVNESVVDNYLYGDLAIDNTRDFNPKLNVKD 484
 Db 255 KIYVRNPEFYLGKISDKGGFNLWNRVNESVVDNYLYGDLAIDNTRDFNPKLNVKD 314
 Query 485 MDWGMKDYKANGPDKVTDMDGNYLQGYSQNLNAKAVGVHQLYDNVKPBNIDPKGN 544
 Db 315 MDWGMKDYKANGPDKVTDMDGNYLQGYSQNLNAKAVGVHQLYDNVKPBNIDPKGN 374
 Query 545 TSIYADGKSVFVNINDKRNNGPGEIQEHIYINGKEYTSFNDIKOIDTKLNK 604
 Db 375 TSIYADGKSVFVNINDKRNNGPGEIQEHIYINGKEYTSFNDIKOIDTKLNK 434
 Query 605 DFARNNTTKEFLNKDGEVSELKPHRTVTIONKEMSTIVSEEDFILPVYKELEG 664
 Db 435 DFARNNTTKEFLNKDGEVSELKPHRTVTIONKEMSTIVSEEDFILPVYKELEG 494
 Query 665 YQFDGWEIISGPGKDKDGYVNLSKDPTIKPVPKKLEKKEBNKPTFDVSKKDNPQVN 724
 Db 495 YQFDGWEIISGPGKDKDGYVNLSKDPTIKPVPKKLEKKEBNKPTFDVSKKDNPQVN 554
 Query 725 HSQLNESHRKELOREHSQKSDSTKDVATVLDKNNISSTNNPK 773
 Db 555 HSQLNESHRKELOREHSQKSDSTKDVATVLDKNNISSTNNPK 603

OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/961,083
 FILING DATE:
 CLASSIFICATION: 435
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Arinieillo, Pamela Deneke
 REGISTRATION NUMBER: 40, 489
 REFERENCE/DOCKET NUMBER: GRC-011
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (781)893-5007
 FAX: (781)893-8277
 INFORMATION FOR SEQ ID NO: 4771:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1529 amino acids
 TYPE: amino acid
 TOPOLogy: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: *Streptococcus pneumoniae*
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (B) LOCATION 1..1529
 SEQUENCE DESCRIPTION: SEQ ID NO: 4771:
 US-09-107-433-4771

Query Match 18.6%; Score 750.5; DB 4; length 1529;
 Best Local Similarity 79.7%; Pred. No. 1.8e-40;
 Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;

QY 1 KUGEIAESKFKNGLNGKEGSLKKDTGVEHHQNEENESIKEKGSKFTIDRNNTIERTDOPENK 60
 Db 1336 KUGBEPESKFKOLKVNKKDLSNKETAVENNLVNDQNSIEGSKLNFNHTKSTIRDPENK 1395

QY 61 DLKLLKKRKFREYDFFSETGKMEEDYKDKGKNTIAYDGTDLKETYETSKLDEIKSI 120
 Db 1396 DLKLLKKRKFQEDDFVN-GGTTVERDYKDKGKNTIAYDGTDLKETYETSKLDEIKSI 1454

QY 121 YGVLSPSKDKHPIBLGKISNVSKNAKTYGGNNVSKIBIKATKVDHFKTMMFDLYANIND 180
 Db 1455 YGVLSPSKDKHPIBLGKISNVSKNAKTYGGNNVSKIBIKATKVDHFKTMMFDLYANIND 1514

QY 181 IVDGLAP 187
 Db 1515 IVDGISF 1521

RESULT 4
 US-08-961-083-68
 Sequence 68, Application US/08961083
 Patent No. 6159469
 GENERAL INFORMATION:
 APPLICANT: Choi et. al.
 TITLE OF INVENTION: *Streptococcus pneumoniae Antigens and Vaccines*
 NUMBER OF SEQUENCES: 452
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/536,784
 FILING DATE: 30-OCT-1997
 CLASSIFICATION: <Unknown>
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: 08/961,083
 FILING DATE: OCT-30-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Michelle S. Marks
 REGISTRATION NUMBER: 41,971
 REFERENCE/DOCKET NUMBER: PB340P3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8512
 FAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 68:

SEQUENCE CHARACTERISTICS:
 LENGTH: 117 amino acids
 TYPE: amino acid
 STRANDBNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 68:
 US-09-536-784-68

Query Match 15.3%; Score 615; DB 4; Length 117;
 Best Local Similarity 100.0%; Pred No. 3.9e-33;
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELKGYOPDGWEISGPEKKDAGYVINSKDTFPIVKKIEKKERENKPTFDVSK 716
 1 YKGELKGYOPDGWEISGPEKKDAGYVINSKDTFPIVKKIEKKERENKPTFDVSK 60

QY 717 KKDNPQHNSQLNESKRKEDIQREENSQKSISTKDTATWDKNNSSKSTNNPK 773
 61 KKDNPQHNSQLNESKRKEDIQREENSQKSISTKDTATWDKNNSSKSTNNPK 117

RESULT 6
 US-09-200-650B-7

Sequence 7, Application US/09200650E
 General Information:
 Applicant: Patti, Joseph M.
 Applicant: Foster, Timothy J.
 Applicant: Hook, Magnus A.O.
 Applicant: Ebdahn, Beidre Ni.
 Applicant: Perkins, Samuel L.
 Title of Invention: Extracellular Matrix Binding Proteins from *Staphylococcus aureus*
 File Reference: P06283US2/BAS
 Current Application: US/09/200,650B
 Current Filing Date: 1998-11-25
 Prior Application Number: 60/066,815
 Prior Filing Date: 1997-11-26
 Prior Application Number: 60/098,427
 Prior Filing Date: 1998-08-31
 Number of SEQ ID NOS: 23
 Software: Patentin Ver. 2.0
 SEQ ID NO 7
 LENGTH: 1166
 TYPE: PRT
 ORGANISM: *Staphylococcus aureus*
 US-09-650B-7

Query Match 5.0%; Score 203; DB 4; Length 1166;
 Best Local Similarity 21.6%; Pred. No. 6.8e-05; Gaps 53;
 Matches 184; Conservative 120; Mismatches 291; Indels 258; Gaps 53;

QY 33 QNEEESTKE-KSSFTIDRNSTIRDPEPKDQKLUKLKKEF-----EVDDFTS 78
 226 KNPCKELKVRNDSTHDKPKVAPTPSKPKVNAKAEPAVOPAANVNDLIK 285

QY 79 FGKRMEEYDKYDDKGKNIYDGGDLEYTE-----KLDEJKSKYGVLS 125.
 286 VTKQTIR---VGDGKDVAAHGDKDIEYDFTIDNKVKRGDTWINTYDKNVPSDLT 341

QY 126 PSKD-----GHEFLGKISVNSKNAK---VYKGNYKSEIKATKVKHFSK-----168
 342 DKNDPDIRDPSGEVIAKGTIDKATKQIYTFTDVKYEDIKSRTLVSIDKKTVPNE 401

QY 169 ---TWT-----DLYANIN-DIVDGIAFAGD--MRLFVK-DNDQKKAETKIRM-PEKIK 214
 402 TSNLNTPATAGKETSONVIVYQDPWIK-GDSNISIFTKIDEDKOTIEQIYVPLK 460

QY 215 ETKSEY---PTVSSYGNVNLGEQGD--LSKKNPKDMLTM-----ESGKLYSEKOO 261
 461 AYNTKVDIAGSQVDDYK-IKLNGNSTIDQTEIKVYQKVNQDQLPQSNRHYDFSOYED 519

QY 262 YLLK-DNLKRGIAKVTTPKGKIDMLENGVSKEDIAKIQKANPMLAISSTIVIA 320

RESULT 7
 US-08-973-462-8

Sequence 8, Application US/08973462B
 General Information:
 Applicant: DRUILHE, PIERRE
 Applicant: DAURILHE, PIERRE
 Title of Invention: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
 File Reference: 0660-0125-0 PCT
 Current Application Number: US/08/973,462B
 Current Filing Date: 1998-02-05
 Earlier Application Number: PCT/FR96/00894
 Earlier Filing Date: 1996-06-12
 Earlier Application Number: FR 95/07007
 Earlier Filing Date: 1995-06-13
 Number of SEQ ID NOS: 29
 SEQ ID NO 8
 LENGTH: 1786
 TYPE: PRT
 ORGANISM: Artificial Sequence
 Feature:
 Other Information: Description of Artificial Sequence: Polypeptide
 US-08-973-462-8

Query Match 5.0%; Score 200.5; DB 3; Length 1786;
 Best Local Similarity 20.5%; Pred. No. 0.00018; Mismatches 329; Indels 301; Gaps 52;
 Matches 204; Conservative 160; Mismatches 329; Indels 301; Gaps 52;

QY 6 AES--KFKN-LGNGKEGLSKKD-----TGVHHHOOE-----ESIK 40
 Db 868 AESVTFNSNLERIOENTITNDTEEKEELHENWLSAALENTQSEEKEKVIDVIEVK 927

41 EKSFSTI-----DRNSTIRD-FEN-----KDLKLUKKKRFEDDPTS 78
 FILE REFERENCE: GTC99-03PA
 CURRENT APPLICATION NUMBER: US/09/328, 352
 Best Local Similarity 19.5%; Pred. No. 0.011; Length 2504;
 Matches 198; Conservative 130; Mismatches 346; Indels 341; Gaps 54;
 SEQ ID NO 5821
 SEQ ID NO 8252
 LENGTH: 2504
 TYPE: PRT
 ; ORGANISM: *Acinetobacter baumannii*
 ; US-09-328-352-5821
 Query Match 4.8%; Score 191.5; DB 4; Length 2504;
 Best Local Similarity 19.5%; Pred. No. 0.011; Length 2504;
 Matches 198; Conservative 130; Mismatches 346; Indels 341; Gaps 54;
 QY 5 IAEFKPKLNGKGEGSLKDKDTGCVHHQNEESEIKEKSFTID-----RNSTIRD 58
 Db 1365 LKOSKFDSV--TTGNTTINTNGL-----TKEGPSITKEGINAGSKQINTVADGT 1412
 QY 59 NDKLKKKKFKEVDDTSETGKRMEEYDVKDD-----KGNIIADGDTLE-- 107
 Db 1413 N-AKDAINKS-QLDNLAAQ-NATDAAKVDDAKTKDVKVTKGK----DGTVDLN 1462
 QY 108 -----YETEK-----LDEBKSKYI--VLSP-----SKDGHFELGK 137
 Db 1463 KAGHSISTSKEAVNGSQHKNSIKNSIGNTVNPDSGLTNNIGGCKNNDAISE 1522
 QY 138 ISNVSKNAK--VYGNNTYKSIETK-----ATKDFHSK-----TMFDLYANINDI 181
 Db 1523 VQNTATAKXTTVEGDN--IVVKEVNKGSTNTYEAVTKDLTNSVTGDTVANNGL 1579
 QY 182 V-----DGLAFAGDMRLFVKNDQKKAECIKRMPKKKETKSEYYVSSIGNVI-- 230
 Db 1580 TIKGSPSTKDGINSAGKQINTVADGINAKDNVQDQTKVKENIN--GRTDTNQND 1637
 QY 231 --ELGEGDLSKQPKPDNLTKMEEGKTYSEKQOYLKDNITLTKGALKVTTNPKTD 287
 Db 1638 AKUDLGNQIADTKNLAQKADQGNQITDNTKNTKQD-----TTOINDTKE 1688
 QY 288 MLESGNG-----VYSEKDIAKIQQKANPNU-----ALSETTVA 320
 Db 1689 LNNTIGNTKELNSKIDSTTLENGLNFGAGNSGNDVHRKLGKLNITIGGAJSTPVK 1748
 QY 321 DS-----RNVEDCRSTOSVLMALD-----GENITRYQVFFKNDKGAIKDQG---- 365
 Db 1749 TSGENVITRTQDGQIQLIKOSKDFSVTTGNTTINTNGLTIC--EGISVKEGINAGG 1805
 QY 366 -----NLTUTSS-----KLVFGKD----- 380
 Db 1806 KKTINADGVNAKDAVNQSDLNLAQKQNTDDAVKDDAKTKDVKVTKGDKGTVLDNV 1865
 QY 381 -----DKE-YTGEDKENE-----AIKEDCSMLFDTKPVNLMSKQYFN- 419
 Db 1866 KACHISSTSKEATNGSQHNNISNSKNSIGNTVNPDS--LTNNIG-GTGNINND 1921
 QY 420 -----PSKSNSKTIYVRNPPEVYRKGKSDKGGENWELRNVESVNDVNYIYGDIAIDNTR 471
 Db 1922 AISEVKNTATKAKITVTEGDNIVVKETKDGTNTVYEVSTKKDTLNSVTGDTVLNN- 1980
 QY 472 DFNKLNKTDGIDMDGMKDYKANGFPDKVTDMDGNVYLQTCYSDLNKAVGVMVQFLY 531
 Db 1981 -----GLTIDGPSI---TKD-GVAGGKKTIDVANGVIAQNSKDAVN----GAQVHHS 2028
 QY 532 NVKEPVNIDPKGNTSIEVADGKSVVFNINDKRNNGFDGRIQEOMIYNGKEYSFNDIKQ 591
 Db 2029 SIRNSIG---GNTVWN-PDGSUTTNTGGKONINDAKS---VDEKVNTGNDQ 2079
 QY 592 TDKTNTKIVWDGPARN--TTVKEFILNDTGTSESELKPHRVVTIONG-----KE 641
 Db 2080 ---KGLN-----FGANDQKTTGKAVHKLGD-----TINIVGSDAKTAEDT 2120
 QY 642 MSSTPVS---EDFLPVVKGELRGYOPDGWELSGF-----ECK 678
 Db 2121 SGENTITRTEGDI-----KIMLKDVKDSVNVYQGHVLAQOGLTIGGGTSITVNGINAGG 2176
 ; RESULT 8
 ; Sequence 5821, application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEAR ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

QY 316 TTYADSRNVEDGRSTOSVUMSALDGNIITRYQVFTFRMNDKGEATDKDGHLVTDSSKLV 375

Db 203 ATT-----LNINNIKHPISIVMT-----GLFTDDQAV 231

QY 376 LIG-KDKEYTGERDENKEAATKEDGSMFIFDTPKPNLMSM-DKVNENPSKNIKIVTRPE 432

Db 232 EWPTEBDSVSGCTIDNGALNBEG-----TRAKNPLPLINSSGAPATGNSNNTKRV 285

QY 433 F---YLRKISDKGGFNNBLRVNESVUDVNYLIYGDLDHIDNTRDFNKLNUVQGDIM--D 486

Db 286 EKDSYQCHA1QAGSKN-----VLTNSRELGOA-----LAKTMUDQOLISKE 329

QY 487 WGMKDYKANGFPDKVUTMDG---NVYLQTY---SDIANK---AVGPHYOFPLYDNYRPE 536

Db 330 IQIBPLTRKGFFYALND-DGKUSVENTIONSYFGKSDKUSGELVTAIGTHYQTLSTONPSN 388

QY 537 VNDPKGNTSISYADGSKVVENINDKNGENGEIQRQHYY-----INCKEYTFEND 588

Db 389 IKLNNHHFDNNMMYAGVRFGTGFTDVLIKNRFDKVKGKESVHYRESGALVNWASYKNTD 448

QY 589 I---KOLI-----DKTNIKIKWKDAR----INTVKEILNNDTGESELK 628

Db 449 LUDLNKQVVIENIENIADPKPKAIR-VAKDABYLQKVSIITVTKVUNINNSKETPO- 505

QY 629 PRHTVTUQNGKMSSTV---SBDFFLPVYKGLB---EKGYQFDWEI---SGFEGK 678

Db 506 PNEBLRVSNDLUVSENSIPIGGSKEGIVIEDSKGKITYLANNQYFLNSKYSKISFIKSANGK 565

QY 679 K-----DAGYVILSKDTPKEPVFKLKEKEEE-----NKPTFDVK 752

Db 566 EPVIRDSIDGNFNT-VTENGLYKLVTNNLSDKNEKEKEKEKOYNSNNVVIDSNQKNGEFNS 624

QY 717 KDNPQVNHQSOLNEHHRK-----EDLQREEHQSKSDTKDV 752

Db 625 SKDNRQMDKDKNQDNCTEENYKIVGDRETTENHNTKSKLIVD 670

RESULT 11

US-08-728-470-10

; Sequence 10, Application US/08728470

; Patent No. 5928651

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: Of No. 5928651-Typeable Haemophilus

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/728,470

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832

FILING DATE: 16-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1

RESULT 12

ATTORNEY/AGENT INFORMATION:

NAME: Berkstresser, Jerry W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-633

TELECOMMUNICATION INFORMATION:

TELEFAX: (703) 415-0810

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1529 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-728-470-10

Query Match 4.7%; Score 187.5; DB 2; Length 1529; Best Local Similarity 20.1%; Pred. No. 0.001; Mismatches 255; Indels 239; Gaps 37;

Matches 148; Conservative 94; Mismatches 255; Indels 239; Gaps 37;

Db 531 RPNNSVINGTK-----GLKFLANQNPFTKETGELANIS-GVTINGNTKVKW 580

QY 64 KLUKKKREVDFTSETGRKRMVEYDY-KYDKGNITAYDDCNDLBYETTEKLEDEKSIYG 122

Db 581 NASKDSYWNVNSLTLNT---VOKFTPKFVDSGS-----NSQDLRSSRRSPAG 625

QY 123 VLSPSKGHFR-ILGKIS-NUSKNAYTY-----GNNYKSI-----KATYDHSKT 169

Db 626 V-----HFNGIGKGNFNGANAKLFLKLPNAATDPKELPITPPNATGNSDSS 678

QY 170 MTFDLVANIND---IVDGLAFAGDMRLFVKONDOKKAETKIRMPPEKIKETKSEPPVS 224

Db 679 WMDIHANTSAAGINMDNSINTGDFDSFTSHNNNSNABIKOLTINATGSNS-- 735

QY 225 SYGVNVIILGEEDLSKONPKDNPNTKMEGKSIYDSEKOQYLLKONTILRGYALKVTTNPG 284

Db 736 -----LKTQKSFNBYSKHAINSSHNL----- 760

QY 285 KTDMLKENGVVSKEDEAKIQKANPNALESETTYADSRNTEGRTOSVUMSALDGFI 344

Db 810 -----LGCVNTLGGNSSSSTGNTNKTQDTSNTNTGKRTLTG----- 841

QY 345 IRYQVFTFKMDKGEAIDKGSNLVTDSSKLVLFGK---DDKEYGDKENVEAKEDGS 400

Db 842 -----ASDNLNITGTNTGNTAN-INKQGVVKLQODDINNGGLN-ITNASGQKTI 892

QY 461 IYGLDHLNTDTRPNKLNK-DGDIMWG-----MDYKAN-----G 496

Db 893 IYGLDHLNTDTRPNKLNK-NIKADAEGIQQGNTISQEGNLTSSDKVNTNQITIKAGVEG 950

QY 497 FPDKVTDMDGNVYLT-----GY-----SDL-----NAKAVGHY 526

Db 951 RDSSEBANANTIQTEKLGDNLSGENKABITAKNGSDLTGNSAGGNAKRY-- 1008

QY 527 QFLYDNTRPEVNIDPKNTSISYADGSKVVENINDKNGNED-GETQEHYINGSKYT 584

Db 1009 --TFDKK-----DSKIST-----DGHNTVNSEVKTNSGNSNAGDNSTGTLISAKDVT 1056

QY 585 SFDIKOIKDITLNIKIKWKFARNTVKE-FILNKDGTGESELKFRVUTQING----- 639

Db 1057 VANNVTS--HKTINISAA---AGNTVKEGTINATITGSV-----VIAQNGIKG 1102

QY 640 --KEMSSSTIVSRDFI 653

Db 1103 NITSQNTVTTAENLV 1118

US-08-719-641-10
; Sequence 10, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REFERENCE/DOCKET NUMBER: 22,651
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-719-641-10
; Query Match Best Local Similarity 4.7%; Score 187.5; DB 3; Length 1529; Matches 148; Conservative 94; Mismatches 255; Indels 239; Gaps 37; QY 9 KPKNL--GNKEGSKKKDTGVEHHQHQENRSIERSIKKSSFTIDRNISTIRDPENKDLK-- 63 Db 531 PRNNVSTNGTQK--GLKFTANQNNPHTKFDGELNIS-GIVTINGTTKKDVKYW 580 Qy 64 KLIKKEFREVDTFSERGKRMKVEYD-KYDKGKNIAYDDGTDLEYETEKUDBIKSKIG 122 Db 581 NASKDSYWNVNSLTNT--VQKPTTRKFVSGS-----NSQDRSSRSFAG 625 Qy 123 VLSPSKGHFE-ILGKTS-NVSKNAKVVY-----GNNYKIEI---KATKVDHSKT 169 Db 626 V-----HENGIGKKNFNGANAKLKFKPNAATDPKELPFPNANTATGNSDS 678 Qy 170 MFDLYANIND----IVDGIAFAGDMRLFVNDOKKAEKIRMPKFKIKTKSEPYVS 224 Db 679 WMDFIHANLTSAAGINMDSINITGGLDFSTISHNRNSNAPBIKKOLITMAGSNSF-- 735 Qy 225 SYGNVIELGEGDLSKKNPDRNITKMSKSYDSEKQOYLDSNLLRKGVALKVITYNG 284 Db 736 -LQTKDUSPFWYNSKHAISNSHLI----- 760

QY 285 KNDMLLGNGVSKEDAKTOKANPMLRALSSTTIVADSRNEDGSTQSIVMSALDGFNI 344 Db 761 ---LGAVNTLGGENSSSITGNNININKANVTLQADTSNTGLKRTLTG----- 809 Qy 345 IRYQVFTRKMDKGAEIKDQNLVTDSSKULFGK---DKEYGDBKENVEATKEDGS 400 Db 810 -----NSVEGLSLTGANANIVGNLSTAEDSTFKE----- 841 Qy 401 MLPIDTPVNLMSMDKIVFPNSKSNKIVYRVPFYLRGKISDKGGFWELRVNESVVDNYL 460 Db 842 -----ASDNINITGFTTNGTAN-INKQGCVKUQGDINNGGGN-ITTNASGQTKI 892 Db 461 YGDLHIDNTRDPNIKANVK-DGD-IMDG-----GY----- 496 Qy 893 INGNI-TNEKGDLNIK-NIKADEI-QIGGNIQSKEGMLTSSDKNINTQITKAGVEGG 950 Db 497 FPDKVUTMDGNGVYLO----- 526 Db 951 RSDSSEANANITIONKRLKAGDINTSGFNAETAAKNGSDLTIGNASGGNADAKV-- 1008 Qy 527 QFLYDVKPKPEVNIDPKGNTSEYADGKSVFVNINDKRNNGFD--GRIQEHOIYINQKEYT 584 Db 1009 --TFDKTK----DSKST----DCHNVTNLSEVTKINGSSNAGNDNSGLTISAKDVT 1056 Qy 585 SPNDIQLQIDKTMNIKVVKDOPARNVTKS-FILNQDTSSELKPHRVTVIIONG--- 639 Db 1057 VNNNTVS--HKINTNSA---AGNTKKGTTINATGSV-----VTAQNGTIG 1102 Qy 640 --KEMSTATIVEEDFTI 653 Db 1103 NITSQWNTVATENLV 1118

RESULT 13
US-09-200-650E-3
; Sequence 3, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.C.
; APPLICANT: Eidhinn, Deirdre Ni
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from *Staphylococcus aureus*
; FILE REFERENCE: P06283U2/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 930
; TYPE: PRT
; ORGANISM: *staphylococcus aureus*
; US-09-200-650E-3
; Query Match Best Local Similarity 4.6%; Score 186.5; DB 4; Length 930; Matches 175; Conservative 121; Mismatches 346; Indels 199; Gaps 42; Qy 3 GBAEKKFKLNGNGKESLKK-DTIGVEHHQ---ENESIKKSSFTIDRNISTIRDFE 58 Db 56 GELNQSKNETTAPSENKTTKVKDQKDNQTTADQPKVTSMSAATVKEETSSMSPQ 115 Qy 59 NDKLKLIKKEFREVDTFSERGKRMKVEYD-KYDKGKNIAYDDGTDLEYETEKUD 117 Db 116 NATANOSTIKTSKNTVNDKSSITYNE----TDKSNLFOAKDWTSTPKTTKPRTLN 169 Qy 118 SKIYGTLSPSKDGHFELGKISNVSKNAKTYGNYKSYEIKATKVDHFSKTMWTDLYAN 177

Db 170 RMAVNTVAAPQQG-----TNV--NDKVHFSN---IDIAIDKGHVQNTGKTERWAT 215 ;
 Qy 178 INDI-----VDSLAFAGDRFLFVKNDPQKAKTAKRMPKIKERKSKPYVSSYGNV 229 ;
 Db 216 SSDVULKUKNYTTIDSVKRGD--TFTFRXGQFRPGSVRLP---SQRONLY--NAQGNT 267 ;
 Qy 230 ISLGEGLDLSKNEP-----DNLTKMSESKYISDSBKKQQYLKDNILRKGYALKVTT 281 ;
 Db 268 TAKGIVYUPTNTTYYTFTNYVQYNTRG----SPEQVATAK----RNATDKTAY 316 ;
 Qy 282 INPGKTDMLEGGNGVSKEDIAKQKQANPMLRASETTYAD----SRVW----EDCRST 331 ;
 Db 317 --KMETLGLNDTYSEETI--UDYGNKKAAQPLISSTVNNINDELSRNTAYNQPKVTT 371 ;
 Qy 332 QSVTMSALDGF---NIRYQVFTKMDKGRAIDKGNLWTDSSKLULRGKDKEVTG 387 ;
 Db 372 KOTFVUTLNTGKPKPNPKNAQKFL--YEVFTDONGFVD--SFTPDTSKL----KDVTF 417 ;
 Qy 388 DKENVEIKEQDSMLFDTKPPNLSMOKNF-----NPSKSNKIVYRNPFYLRGK 439 ;
 Db 418 DDFDV--IYSNDNKTATVDLMLKGOTSSNKOYIICQAVYPDNNSSTDN-----GKI 464 ;
 Qy 440 -----SDKGKGFNWE--LRVNTSVVDNLYLXGDLHIDNTRDENIKLUVKGD--DIMWGM 489 ;
 Db 465 DYLDTDKTIVKYSNSNSYNSSTAN--GDQKKNLGLDVWEDTNDKQDQDANBKG 520 ;
 Qy 490 KQY----KANGFP--DKVTDMDGCVYVLYOTGYSIDNARAKAVGHL-----YQFLYDNK 534 ;
 Db 521 KGVVILKDSNGKELDRDITDENGKQOPTGLSN-----GTYISVEFSTPAGYPTPTANV 574 ;
 Qy 535 PEVNIDPKGNTS--IYADG-----KSVUNINMK--RNNFEGDEIQRPHIYNGK 581 ;
 Db 575 TDDAVDSDGLTTRGIVKODAAMTLDGSFYKTPKSYLDYVWDSNKGKRDSTEGKIKGV 634 ;
 Qy 582 EYTSFNDTQKQDIDTLLNKKVDFARNTTVEKFLNKTGESEL--KPHRTVTION- 638 ;
 Db 635 KVTQLNQKGEVIGTT-----EDENGKQRFDULSGKVKIFERKPAQLTQGNT 684 ;
 Qy 639 -----GKMSSTIVSEEDFLILPVYKGELESKGYQDFGWRISGFECKKDACYVNLSKD 690 ;
 Db 685 TEDDKDADGGEVDVTIDHDF-----TLONGYEEETSDSDSDSDSDSDSDSDSD 737 ;
 Qy 691 TFKPVVKIEKKEERKMPTEVSKCKDNPQVNHSQNLNEHSRKEDIQREHNSQSKSTK 750 ;
 Db 738 S-----DSD 788 ;
 Qy 751 D 751 ;
 Db 789 D 789 ;

RESULT 14
 US-09-200-650E-5
 ; Sequence 5, Application US/09200650E
 ; Patent No. 6690195
 ; GENERAL INFORMATION:
 ; APPLICANT: Patti, Joseph M.
 ; APPLICANT: Foster, Timothy J.
 ; APPLICANT: Hock, Magnus A.O.
 ; APPLICANT: Eichman, Deirdre Ni
 ; APPLICANT: Perkins, Samuel L.
 ; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from *Staphylococcus aureus*
 ; FILE REFERENCE: P0283052/BAS
 ; CURRENT APPLICATION NUMBER: US/09/200,650E
 ; CURRENT FILING DATE: 1998-11-25
 ; PRIORITY APPLICATION NUMBER: 60/066,815
 ; PRIORITY FILING DATE: 1997-11-26
 ; PRIORITY NUMBER: 60/098,427
 ; NUMBER OF SEQ ID NOS: 23
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 5
 ; LENGTH: 1315

Db 494 ANGFPK-----VTDMDGNTVLYQGYSIDNAKAVVHYQI--YDNVKPEVNID 540 ;
 Qy 541 PKGNTSTIEYADGKSVVNFINDKRN---NGDGE10EQHIVINGKFTSFNDIKQIDKT 596 ;
 Db 541 PKGNTSTIEYADGKSVVNFINDKRN---NGDGE10EQHIVINGKFTSFNDIKQIDKT 596 ;
 Db 748 PEGYTPPTVSGSDIEKDSNGLTTCVINGADNMFLDSGFYKTPK--YNLGNVWEDT 803 ;
 Qy 597 LNIKIVVDFARNTTVEKFLNKTGESELKP--HRVVTIONGKEMSTIVSEEDFLP 655 ;
 Db 804 -----NKDGKDPSTEKGKISGVWVTLKN-----ENGEVLO 832 ;
 Qy 656 VTKGELLKGYOPDGWESGF--EGKIDAGYVNLSTKTFIPIKVKKIEBKEBENKPTD 713 ;
 Db 833 TTKTDKQKGYOPTGLENGTGYKVEFETPSGYPT----QVGSGTDEGIDSNGSTTG 884 ;
 Qy 714 VSKKDKDQPVNHSQNLNEHSRKEDIQREHNSQSKSTK-----VPAVLDKNNISKST 767 ;
 Db 885 VTKDKDNTIDSFGYKPTYNGDYYWEDTNTKNGVQDKEKISGVWVTLKODENDVLT 944 ;
 Qy 768 TNPN 772 ;
 Db 945 TTDEN 949 ;

US-09-949-016-11433
; Sequence 11433, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; TIME OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastaSEQ for Windows Version 4.0

QY 593 IDKDTNLTKIVVQDFARNNTVVKERTLINKGTVSEILK--PHRTVTTIONKGKEMSTIVSRB 650
Db 1909 VDKTNQLMBTLTKTICKNTOQKAQLDSFVKSMSLQNDRDRITVGDYQOLEBERHSLTIEK 1968
QY 651 DFILPVYKGELKEQYQFDQWEISGFGEG-----KKDAGYVILNSKDT 691
Db 1969 DQLIQEAAANENLKE---EIRGLRSHMDLNSENAQLDAELIQYREDLNQVITI-KDS 2023
QY 692 FIKRUVFK-KIEKEKEENKPTFDVSKKCDNPQYHQSOLNES---HRKEDIQREERSOK 745
Db 2024 QKQKOLLEVOLOQNKELNKYAKLUEKLUKESEAN-EDLRRSFINALQEEBKQDLSKEIESLK 2082
QY 746 --SDSTKUVA 754
Db 2083 VTSQLTQVTA 2094

Search completed: February 16, 2005, 19:20:41
Job time : 50 Becs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM Protein - protein search, using SW model

Run on: February 16, 2005, 19:19:57 ; Search time 141 Seconds

(without alignments)
1791.327 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGEBIAESKPKNLGNGKEGS.....ATVLDKNNISKSTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptcdata/2/pupaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptcdata/2/pupaa/PCT NEW_PUB.pep:*

3: /cgn2_6/ptcdata/2/pupaa/US06_NEW_PUBCOMB.pep:*

4: /cgn2_6/ptcdata/2/pupaa/US05_NEW_PUB.pep:*

5: /cgn2_6/ptcdata/2/pupaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptcdata/2/pupaa/PCTUS_PUBCOMB.pep:*

7: /cgn2_6/ptcdata/2/pupaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptcdata/2/pupaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptcdata/2/pupaa/US09A_PUBCOMB.pep:*

10: /cgn2_6/ptcdata/2/pupaa/US09B_PUBCOMB.pep:*

11: /cgn2_6/ptcdata/2/pupaa/US09C_PUBCOMB.pep:*

12: /cgn2_6/ptcdata/2/pupaa/US09_NEW_PUB.pep:*

13: /cgn2_6/ptcdata/2/pupaa/US10_PUBCOMB.pep:*

14: /cgn2_6/ptcdata/2/pupaa/US10B_PUBCOMB.pep:*

15: /cgn2_6/ptcdata/2/pupaa/US10C_PUBCOMB.pep:*

16: /cgn2_6/ptcdata/2/pupaa/US10D_PUBCOMB.pep:*

17: /cgn2_6/ptcdata/2/pupaa/US10D_NEW_PUB.pep:*

18: /cgn2_6/ptcdata/2/pupaa/US11_NEW_PUB.pep:*

19: /cgn2_6/ptcdata/2/pupaa/US60_NEW_PUB.pep:*

20: /cgn2_6/ptcdata/2/pupaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Query Match Length DB ID Description

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	773	13	US-10-067-385-8
2	4026	100.0	2119	10	US-09-769-744A-28
3	4026	100.0	2140	15	US-10-882-122A-7670
4	4026	100.0	2140	17	US-10-472-928-1180
5	615	15.3	117	9	US-09-765-72-68
6	56	11.96	15	US-10-882-122A-52737	
7	222	5.5	861	10	US-09-820-843A-109
8	218	5.4	1639	14	US-10-087-464-10
9	212	5.3	1441	15	US-10-1082-122A-70251
10	207.5	5.2	4688	15	US-10-282-122A-76865
11	5.1	1191	15	US-10-282-122A-52048	
12	203	5.0	1186	17	US-10-144-616-7
13	5.0	1881	14	US-10-032-585-7646	

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0%	773	13	US-10-067-385-8
2	4026	100.0%	2119	10	US-09-769-744A-28
3	4026	100.0%	2140	15	US-10-882-122A-7670
4	4026	100.0%	2140	17	US-10-472-928-1180
5	615	15.3	117	9	US-09-765-72-68
6	56	11.96	15	US-10-882-122A-52737	
7	222	5.5	861	10	US-09-820-843A-109
8	218	5.4	1639	14	US-10-087-464-10
9	212	5.3	1441	15	US-10-1082-122A-70251
10	207.5	5.2	4688	15	US-10-282-122A-76865
11	5.1	1191	15	US-10-282-122A-52048	
12	203	5.0	1186	17	US-10-144-616-7
13	5.0	1881	14	US-10-032-585-7646	

Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 121 YGULSPSKDGHFELIGKISNSKNAKYIGNYKSEIKATKDFASFKTMFDLYANIND 180
 Qy 181 IVDGLAFAGDMLFLVKONDOKKAIEKIRMPBKETKSEPYVSSYGNVILGEGLSKN 240
 Db 181 IVDGLAFAGDMLFLVKONDOKKAIEKIRMPBKETKSEPYVSSYGNVILGEGLSKN 240
 Qy 241 KPDNLTKMESEKLYSEKQYLKONILIKGYALKVTTNPCKTMDLNGVYSKEDI 300
 Db 241 KPDNLTKMESEKLYSEKQYLKONILIKGYALKVTTNPCKTMDLNGVYSKEDI 300
 Qy 301 AKIOKANPNALESETTYIASRNVEDGRSOSVMSALDFNIRIQVFTKMDKGEA 360
 Db 301 AKIOKANPNALESETTYIASRNVEDGRSOSVMSALDFNIRIQVFTKMDKGEA 360
 Db 361 IDKGNLVTDSKLUFLGKDKDKEYTECDKFNTEAKEDGSMFLIDTKPVNLSMDKNYFNP 420
 Db 361 IDKGNLVTDSKLUFLGKDKDKEYTECDKFNTEAKEDGSMFLIDTKPVNLSMDKNYFNP 420
 Qy 421 SKSNKTYVRNPFYFLRGKISDKGGFWNLRNVESVUDNLYIYGDHIDNTRDFNKLNVK 480
 Db 421 SKSNKTYVRNPFYFLRGKISDKGGFWNLRNVESVUDNLYIYGDHIDNTRDFNKLNVK 480
 Qy 481 DGDIMDGKMDYKANGFPDKTMDGNGVYIQLGYSIDLNAKAVGHYQFLDNVKEVNID 540
 Db 481 DGDIMDGKMDYKANGFPDKTMDGNGVYIQLGYSIDLNAKAVGHYQFLDNVKEVNID 540
 Qy 541 PKGNTSEYADGSKVWNINDKRNGFDGETQEOHYINGKEYTSFNDIKOIDKLNIK 600
 Db 541 PKGNTSEYADGSKVWNINDKRNGFDGETQEOHYINGKEYTSFNDIKOIDKLNIK 600
 Qy 661 LRGKYOPDGWEISGFEGKDKAGYVINSKOTPIKPFKKIEKEKENPKFDVSKDN 720
 Db 661 LRGKYOPDGWEISGFEGKDKAGYVINSKOTPIKPFKKIEKEKENPKFDVSKDN 720
 Qy 721 P0VNH5OLNE5HRSKEDLQREBHSQKSDSTKDVATVLDKONISSKTTNNPK 773
 Db 721 P0VNH5OLNE5HRSKEDLQREBHSQKSDSTKDVATVLDKONISSKTTNNPK 773

RESULT 2
 US-00-769-744A-28
 ; Sequence 28, Application US/09769744A
 ; Publication No. US20030134407A1
 GENERAL INFORMATION:
 APPLICANT: Le Page, Richard WP
 APPLICANT: Wells, Jeremy M
 APPLICANT: Hanniffy, Sean B
 APPLICANT: Hansbro, Phillip M
 TITLE OF INVENTION: Proteins
 FILE REFERENCE: PNC/P21122WO
 CURRENT APPLICATION NUMBER: US/09/769, 744A
 PRIOR APPLICATION NUMBER: PCT/GB99/02452
 PRIOR FILING DATE: 1999-07-27
 PRIOR APPLICATION NUMBER: GB 9836336 .3
 PRIOR FILING DATE: 1998-07-27
 PRIOR APPLICATION NUMBER: US 60/125329
 PRIOR FILING DATE: 1999-03-19
 NUMBER OF SEQ ID NOS: 196
 SOFTWARE: Patentin Ver. 2.1
 SEQ ID NO: 28
 LENGTH: 2119
 TYPE: PRT
 ORGANISM: Streptococcus pneumoniae
 US-09-769-744A-28

Query Match Local Similarity 100.0%; Score 4026; DB 10; Length 2119;
 Best Local Similarity 100.0%; Pred. No. 5.8e-216; Length 2119;

Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1313 KLGELAESFKNGLNGKEGSLSKKTGTGVERHQAENBESIKKSSTIDRNISTRDFENK 60
 Qy 61 DKKLKKERFEDPTEGKMRMERYDKDKNIATDGTLEYETEKLDEIKSI 120
 Db 1373 DKKLKKERFEDPTEGKMRMERYDKDKNIATDGTLEYETEKLDEIKSI 1432
 Qy 121 YGULSPSKDGHFELIGKISNSKNAKYIGNYKSEIKATKDFHSKMTFDLYANIND 180
 Db 1433 YGULSPSKDGHFELIGKISNSKNAKYIGNYKSEIKATKDFHSKMTFDLYANIND 1492
 Qy 181 IVDGLAFAGDMLFLVKONDOKKAIEKIRMPBKETKSEPYVSSYGNVILGEGLSKN 240
 Db 1493 IVDGLAFAGDMLFLVKONDOKKAIEKIRMPBKETKSEPYVSSYGNVILGEGLSKN 1552
 Qy 241 KPDNLTKMESEKLYSEKQYLKONILIKGYALKVTTNPCKTMDLNGVYSKEDI 300
 Db 1553 KPDNLTKMESEKLYSEKQYLKONILIKGYALKVTTNPCKTMDLNGVYSKEDI 1612
 Qy 301 AKIOKANPNALESETTYIASRNVEDGRSOSVMSALDFNIRIQVFTKMDKGEA 360
 Db 1613 AKIOKANPNALESETTYIASRNVEDGRSOSVMSALDFNIRIQVFTKMDKGEA 1672
 Qy 421 SKSNKTYVRNPFYFLRGKISDKGGFWNLRNVESVUDNLYIYGDHIDNTRDFNKLNVK 480
 Db 1733 SKSNKTYVRNPFYFLRGKISDKGGFWNLRNVESVUDNLYIYGDHIDNTRDFNKLNVK 1792
 Qy 481 DGDIMDGKMDYKANGFPDKTMDGNGVYIQLGYSIDLNAKAVGHYQFLDNVKEVNID 540
 Db 1793 DGDIMDGKMDYKANGFPDKTMDGNGVYIQLGYSIDLNAKAVGHYQFLDNVKEVNID 1852
 Qy 541 PKGNTSEYADGSKVWNINDKRNGFDGETQEOHYINGKEYTSFNDIKOIDKLNIK 600
 Db 1853 PKGNTSEYADGSKVWNINDKRNGFDGETQEOHYINGKEYTSFNDIKOIDKLNIK 1912
 Qy 601 IVDGLAFAGDMLFLVKONDOKKAIEKIRMPBKETKSEPYVSSYGNVILGEGLSKN 240
 Db 1913 IVDGLAFAGDMLFLVKONDOKKAIEKIRMPBKETKSEPYVSSYGNVILGEGLSKN 240
 Qy 661 LRGKYOPDGWEISGFEGKDKAGYVINSKOTPIKPFKKIEKEKENPKFDVSKDN 720
 Db 1973 LRGKYOPDGWEISGFEGKDKAGYVINSKOTPIKPFKKIEKEKENPKFDVSKDN 2032
 Qy 721 P0VNH5OLNE5HRSKEDLQREBHSQKSDSTKDVATVLDKONISSKTTNNPK 773
 Db 2033 P0VNH5OLNE5HRSKEDLQREBHSQKSDSTKDVATVLDKONISSKTTNNPK 2085

RESULT 3
 US-10-282-122A-73670
 ; Sequence 73670, Application US/10282122A
 ; Publication No. US20040029129A1
 GENERAL INFORMATION:
 APPLICANT: Wang, Liangsu
 APPLICANT: Zamudio, Carlos
 APPLICANT: Malone, Cheryl
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari
 APPLICANT: Zyskind, Judith
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John
 APPLICANT: Carr, Grant
 APPLICANT: Yamamoto, Robert
 APPLICANT: Forstch, R.
 APPLICANT: Xu, H.
 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: EULTRA 034A
 CURRENT APPLICATION NUMBER: US/10/282.122A
 CURRENT FILING DATE: 2003-02-20
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/230,335
 PRIOR FILING DATE: 2000-09-06
 PRIOR APPLICATION NUMBER: 60/230,347
 PRIOR FILING DATE: 2000-09-09
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/267,636
 PRIOR FILING DATE: 2001-02-09
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 Remaining Prior Application data removed - See File Wrapper or PALM.
 NUMBER OF SEQ ID NOS: 78614

SEQ ID NO: 73670
 LENGTH: 2140
 TYPE: BRT
 ORGANISM: Streptococcus pneumoniae
 IS-10-282.122A-73670

Query Match 100.0%; Score 4026; DB 15; length 2140;
 Best Local Similarity 100.0%; Pred. No. 5.9e-216; Mismatches 0; Indels 0; Gaps 0;
 Matches 773; Conservative 0; MisMatches 0; InDelS 0;

1 KUGELIASPKPKVNGKCGSLKDTTGYEHHRQNEESIKEKSFTIDRNISTRDPEK
 1334 KUGELIASPKPKVNGKCGSLKDTTGYEHHRQNEESIKEKSFTIDRNISTRDPEK 60

61 DKKLKKKKFREVDFFSETGKRMEEVYKYDKDGNIAYDDGTDLEYETEKEDEKSKI 120
 1394 DKKLKKKKFREVDFFSETGKRMEEVYKYDKDGNIAYDDGTDLEYETEKEDEKSKI 1453

121 YGVLSPSKDKGHFELGIKLSNVSKNAKVVGNVYKSYIEIKATKDFHSKMTFDLYANIND 180
 1454 YGVLSPSKDKGHFELGIKLSNVSKNAKVVGNVYKSYIEIKATKDFHSKMTFDLYANIND 1513

181 IVDGLAGFAGDMRLFVKNDQKAEIKIRMPKEKETKSEPYVSSYGNVIELGEGLSKN 240
 1514 IVDGLAGFAGDMRLFVKNDQKAEIKIRMPKEKETKSEPYVSSYGNVIELGEGLSKN 1573

241 KPDNLTKMSEKSYLDSBQKQYLLKDTNITRKYALKVTTYNGKTMLEGNGVSYKEDI 300
 1574 KPDNLTKMSEKSYLDSBQKQYLLKDTNITRKYALKVTTYNGKTMLEGNGVSYKEDI 1633

301 AKIQKANPMLRALSETTIYADSRNVEDGRSTQSVLMSALDFNIRYQVTFKMDKGEA 360
 1634 AKIQKANPMLRALSETTIYADSRNVEDGRSTQSVLMSALDFNIRYQVTFKMDKGEA 1693

361 IDKGNLVTDSKVLFLGKDKDKEYTGDEKFNVERAIEKEDGMSMLFTDTPVNLSDKMYNRP 420
 1694 IDKGNLVTDSKVLFLGKDKDKEYTGDEKFNVERAIEKEDGMSMLFTDTPVNLSDKMYNRP 1753

421 SKENKTYNPERYLRSKISDKGGPWNWLVRNSVVDVYLLKDHINTDRDNKLNK 480
 1754 SKENKTYNPERYLRSKISDKGGPWNWLVRNSVVDVYLLKDHINTDRDNKLNK 1813

481 DGDIMPQGMDKQYKANGFDPKVTDGNYLQGYSQDNLAKAVGHYQPLDNYKPNV 540
 1814 DGDIMPQGMDKQYKANGFDPKVTDGNYLQGYSQDNLAKAVGHYQPLDNYKPNV 1873

541 PKGNTSEYADGKSWVFINDKRNGFGEIQSHQHINYKGVTSFNDIKQIQLDTLNK 600

RESULT 4 US-10-472-928-1180
 ; Sequence 1180, Application US/10472928
 ; Publication No. US20050020813A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CHIRON SPA
 ; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
 ; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
 ; FILE REFERENCE: P0262660
 ; CURRENT APPLICATION NUMBER: US/10/472,928
 ; CURRENT FILING DATE: 2003-09-26
 ; PRIOR APPLICATION NUMBER: GB-0107658.7
 ; PRIOR FILING DATE: 2001-03-27
 ; NUMBER OF SEQ ID NOS: 4979
 ; SOFTWARE: SeqWin99, version 1.03
 ; SEQ ID NO: 1180
 ; LENGTH: 2140
 ; TYPE: PRY
 ; ORGANISM: Streptococcus pneumoniae
 ; FEATURE:
 ; OTHER INFORMATION: Berrine protease, subtilase family
 ; OTHER INFORMATION: Cellulase location: peptidoglycan-bound (LPXTG)
 ; OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.E+01)
 ; US-10-472-928-1180

Query Match 100.0%; Score 4026; DB 17; length 2140;
 Best Local Similarity 100.0%; Pred. No. 5.9e-216; Mismatches 0; Indels 0; Gaps 0;
 Matches 773; Conservative 0; MisMatches 0; InDelS 0;

1 KUGELIASPKPKVNGKCGSLKDTTGYEHHRQNEESIKEKSFTIDRNISTRDPEK
 1334 KUGELIASPKPKVNGKCGSLKDTTGYEHHRQNEESIKEKSFTIDRNISTRDPEK 60

61 DKKLKKKKFREVDFFSETGKRMEEVYKYDKDGNIAYDDGTDLEYETEKEDEKSKI 120
 1394 DKKLKKKKFREVDFFSETGKRMEEVYKYDKDGNIAYDDGTDLEYETEKEDEKSKI 1453

121 YGVLSPSKDKGHFELGIKLSNVSKNAKVVGNVYKSYIEIKATKDFHSKMTFDLYANIND 180
 1454 YGVLSPSKDKGHFELGIKLSNVSKNAKVVGNVYKSYIEIKATKDFHSKMTFDLYANIND 1513

181 IVDGLAGFAGDMRLFVKNDQKAEIKIRMPKEKETKSEPYVSSYGNVIELGEGLSKN 240
 1514 IVDGLAGFAGDMRLFVKNDQKAEIKIRMPKEKETKSEPYVSSYGNVIELGEGLSKN 1573

241 KPDNLTKMSEKSYLDSBQKQYLLKDTNITRKYALKVTTYNGKTMLEGNGVSYKEDI 300
 1574 KPDNLTKMSEKSYLDSBQKQYLLKDTNITRKYALKVTTYNGKTMLEGNGVSYKEDI 1633

301 AKIQKANPMLRALSETTIYADSRNVEDGRSTQSVLMSALDFNIRYQVTFKMDKGEA 360
 1634 AKIQKANPMLRALSETTIYADSRNVEDGRSTQSVLMSALDFNIRYQVTFKMDKGEA 1693

361 IDKGNLVTDSKVLFLGKDKDKEYTGDEKFNVERAIEKEDGMSMLFTDTPVNLSDKMYNRP 420
 1514 IDKGNLVTDSKVLFLGKDKDKEYTGDEKFNVERAIEKEDGMSMLFTDTPVNLSDKMYNRP 1753

421 KPDNLTKMSEKSYLDSBQKQYLLKDTNITRKYALKVTTYNGKTMLEGNGVSYKEDI 300
 1574 KPDNLTKMSEKSYLDSBQKQYLLKDTNITRKYALKVTTYNGKTMLEGNGVSYKEDI 1633

301 AKIQKANPMLRALSETTIYADSRNVEDGRSTQSVLMSALDFNIRYQVTFKMDKGEA 360
 1634 AKIQKANPMLRALSETTIYADSRNVEDGRSTQSVLMSALDFNIRYQVTFKMDKGEA 1693

361 IDKGNLVTDSKVLFLGKDKDKEYTGDEKFNVERAIEKEDGMSMLFTDTPVNLSDKMYNRP 420
 1694 IDKGNLVTDSKVLFLGKDKDKEYTGDEKFNVERAIEKEDGMSMLFTDTPVNLSDKMYNRP 1753

QY 421 SSKSNKTYVNPFPYRLRCKSDKGPNWLRKANESVNDYLIGDLIDNTDPNKLVK 480
 Db 1754 SKSNKTYVNPFPYRLRCKSDKGPNWLRKANESVNDYLIGDLIDNTDPNKLVK 1813
 QY 481 DGDIDMPGMKDIXKANGSPDKVTDGMDGNYLPGYSDANAKVGWYQFLYNPKPENID 540
 Db 1814 DGDIDMPGMKDIXKANGSPDKVTDGMDGNYLPGYSDANAKVGWYQFLYNPKPENID 1873
 QY 541 PKGNTSIEYACKSVWNINDRNGDGEQEHQYINGKEYTSNDIKIIDIKLNK 600
 Db 1874 PKGNTSIEYACKSVWNINDRNGDGEQEHQYINGKEYTSNDIKIIDIKLNK 1933
 QY 601 IVKDFARNTTKEFLNKDGEVSEKPHRTVTONGRMSNTVSEERFLPYPGE 660
 Db 1934 IVKDFARNTTKEFLNKDGEVSEKPHRTVTONGRMSNTVSEERFLPYPGE 1993
 QY 661 LRGYQFDGWEISGFBGKDAGYVUNLSKDTIKPVEKKERKEBNKPTFDVSKKD 720
 Db 1994 LRGYQFDGWEISGFBGKDAGYVUNLSKDTIKPVEKKERKEBNKPTFDVSKKD 2053
 QY 721 PQVNHSOLNESHRKEDLQRBSOKSDTQVATVLDKNMISSTNNPK 773
 Db 2054 PQVNHSOLNESHRKEDLQRBSOKSDTQVATVLDKNMISSTNNPK 2106

RESULT 5
 US-03-765-272-68
 ; Sequence 68, Application US/09765272
 ; Patent No. US2002006155A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi, et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20855
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HR Vectra 486/33
 ; OPERATING SYSTEM: MS-DOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; PRIORITY APPLICATION NUMBER: 35,373
 ; PRIORITY DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8512
 ; TELEFAX: (301) 309-8504
 ; INFORMATION FOR SEQ ID NO: 68:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 117 amino acid
 ; TYPE: amino acid
 ; STRANDBENDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
 ; US-03-765-272-68

Query Match 5 6%; Score 227; DB 15; Length 1196;
 Best Local Similarity 20.2%; Pred. No. 0.0029%; Matches 185; Conservative 145; Mismatches 260; Indels 328; Gaps 46;

QY 1 KLGETAEKFQKLGNGKEG-----SIKKITGVHHHQNEESTI 39
 Db 123 RLKDIQE-LFENDTGKKEGGSIIIGGKIEAVLGSKPERRSILLEAGTVKFKWKEED 181
 QY 40 KEKSSFTIDRNISTRDFEN--KDKLKKKKFREVDF--TSEIGKRM----- 85
 Db 182 KKLISN-TEQNLIRKDLILWTEERMEPLKEESERAKKFLNLSLEKKEVNMYSIDK 239
 QY 86 -EVDYKDDKGNTIAYDDGND-LEYETEKUDEIKSKYIGLPSKDGHEFTLGKISNVSK 143
 ; ORGANISM: Clostridium botulinum
 ; US-10-282-122A-52737

Db 240 IEKDLK-NISNMLSLGENIDNIKNEQKEYTISKP-----NEKLELDK--NNSR 288 ; NAME/KEY: misc_feature
; OTHER INFORMATION: hypothetical protein
Qy 144 NAKVYGNNTYKSTEIKAKYDFEISKTWMFDLYANINDIVDGLAFAGIOMRLFTKDQKA 203 ; NAME/KEY: misc_feature
; OTHER INFORMATION: gi|4493994
Db 289 NKEBYYVYNNKDKNKDE-----NENALKEKIKNLD-----NICKYKENTLTKN 331 ; US-09-820-843A-109
Qy 204 EIKR-----MPEKIKETKSEPYV----- 223 ; Query Match 5.5%; Score 222; DB 10; Length 861;
Db 332 EERLOQYKLEKEBALEKKLNKLKEEESYLEREICKKEDVNENFENKELKEKEKULKILSBE 391 ; Best Local Similarity 19.3%; Pred. No. 0_00037; Matches 162;保守性 136; Mismatches 324; Index 216; Gaps 37;
Qy 224 -----SSYGNVIELGEGDL--SKNEDDNLTMEGSKYI-----DSBKKQYILK 265 ; Matches 154 LINNIGKUFKNLDDNCNLTPLNLYDKEG--KMHJTGKETNAASSENEYNHANKTNN 211
Db 392 IEELSNNTSNLRBISWMEIIVULENGLDNKRNCSDSYISININIKEDKEKINIK 451 ; NAME/KEY: misc_feature
Qy 266 DNITL-----RKGYALKVYTNPCKTDMLEGNGTYSKD-----IAKIOK---AN 307 ; OTHER INFORMATION: gi|4493994
Db 452 ENILLENLNKENSRSNSSLKISLNKKK-KLKCKNAAYSRLUEANYHMLNLSLEKHYEGYN 510 ; Query Match 5.5%; Score 222; DB 10; Length 861;
Qy 308 PNURALSETTYIADSRAVVEDGRSTQSVLMSALDGFNIIRYQVFTKNDKGEBADKGNL 367 ; Best Local Similarity 19.3%; Pred. No. 0_00037; Matches 162;保守性 136; Mismatches 324; Index 216; Gaps 37;
Db 511 RSVKTLMHEVSKYKUDNITGG-----CEVIGDIIKTYKELETTAMEITALGGRAI---SNV 560 ; Matches 48 DENISTRDPENKL-----KULJJKKREVDFTSETGKRMEEYDYKDDKNTI 98
Qy 368 VTDSSKVLVLFKGDKDKEYTGEDEKNSMLFIDTKPVNLSMDKMYFNP--SKSN 424 ; Matches 212 NNNSYNNNNFCNNNNNNNNNNNNNNNNKMGKNGKERSLNUYLK--KEHMDVBYNNNGKIR 268
Db 561 ITDE-----NKAKLILINYKK-----SLGRATFPLTIQGR 594 ; Query Match 5.5%; Score 222; DB 10; Length 861;
Db 638 -----MSDALKIAKLUKNYSEFKVITLGEV----- 665 ; Best Local Similarity 19.3%; Pred. No. 0_00037; Matches 162;保守性 136; Mismatches 324; Index 216; Gaps 37;
Qy 425 KIVYRNPNPFYLRGKISDKKGFFNWLRYNESWVNLYTIGDLDHIDNTRDPNT--KLYNTKG 482 ; Matches 99 AYT---DGTDLEYEFTBKLBKSKYVGLSPSKOHPET--LGKISNVNSNAKQYGN 151
Db 595 KAKINN-----VTRBDG---LGIASLDIY---DVKESNIDIVLGLRTVAKD- 637 ; Query Match 5.5%; Score 222; DB 10; Length 861;
Qy 483 DNDWGMKDYKANGEPDKVTMDGNYVLTQGVSDLNKAQWGHYQFLYDNVREVNIDPK 542 ; Best Local Similarity 19.3%; Pred. No. 0_00037; Matches 162;保守性 136; Mismatches 324; Index 216; Gaps 37;
Db 638 -----MSDALKIAKLUKNYSEFKVITLGEV----- 665 ; Matches 269 KNDSEKTYWDNPPLHY-----SKKNYDIFTLGDKKYYAKNNNGKNN 310
Qy 543 GNT---SIEYADGSKSVVNNINDKNGNFDGELQEOHQIYINGKEYTSFNDIKQIDKTLNI 599 ; Query Match 5.5%; Score 222; DB 10; Length 861;
Db 666 GSPJTGGSIKHRAQSSI---ISRK-----RIBE-----TCKBLEETKN- 700 ; Best Local Similarity 19.3%; Pred. No. 0_00037; Matches 162;保守性 136; Mismatches 324; Index 216; Gaps 37;
Db 729 ELI-----YNTNTEBITKFMGKLNKAEDTERLRLSSNLTISREEKUTDKKIDBENTINNSOK 784 ; Matches 397 GNNII-----NNNSNKNNFIQOCNSRNVYVMVNNNNNNNTISNNKEASNTD 447
Qy 701 -----TIEBFGN-----ILENNKNIK---LDEENANI---KD 728 ; Query Match 5.5%; Score 222; DB 10; Length 861;
Db 660 ELKKGYODGWEISGPEGKCKDA-----GVVNLNSKD--TPIKPVFKKTRKEKENK 709 ; Best Local Similarity 19.3%; Pred. No. 0_00037; Matches 162;保守性 136; Mismatches 324; Index 216; Gaps 37;
Db 729 ELI-----YNTNTEBITKFMGKLNKAEDTERLRLSSNLTISREEKUTDKKIDBENTINNSOK 784 ; Matches 397 GNNII-----NNNSNKNNFIQOCNSRNVYVMVNNNNNNNTISNNKEASNTD 447
Qy 710 PTFDVSKKDNQPNHOLSNLESRKEDIQREBNSQSKDSTK---DVTAVLKDNNISSK 765 ; Query Match 5.5%; Score 222; DB 10; Length 861;
Db 785 QLEBLKLKD---LNHNDIKEC---EDFLQNEENVKNDKLUJYEKTEKAKUDEMVLVS 839 ; Best Local Similarity 19.3%; Pred. No. 0_00037; Matches 162;保守性 136; Mismatches 324; Index 216; Gaps 37;
Qy 766 -----STTNPNK 773 ; Query Match 5.5%; Score 222; DB 10; Length 861;
Db 840 KELYSMDNTINLANNENK 857 ; Best Local Similarity 19.3%; Pred. No. 0_00037; Matches 162;保守性 136; Mismatches 324; Index 216; Gaps 37;

RESULT 7 ;
US-09-820-843A-109 ;
; Sequence 109, Application US/09820843A ;
; Publication No. US20030039963A1 ;
; GENERAL INFORMATION: ;
; APPLICANT: Council of Scientific and Industrial Research ;
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN ;
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES ;
; FILE REFERENCE: Q63915 ;
; CURRENT APPLICATION NUMBER: US/09/820,843A ;
; CURRENT FILING DATE: 2001-03-30 ;
; NUMBER OF SEQ ID NOS: 118 ;
; SOFTWARE: PatentIn version 3.0 ;
; SEQ ID NO 109 ;
; LENGTH: 861 ;
; TYPE: PPT ;
; ORGANISM: Plasmodium falciparum ;
; FEATURE: ;

RESULT 8 ;
US-10-087-444-10 ;
; Sequence 10, Application US/10087464 ;
; Publication No. US20030059436A1 ;
; GENERAL INFORMATION: ;
; APPLICANT: Chishti, Athar ;

APPLICANT: Oh, Steven
APPLICANT: Liu, David
APPLICANT: Goel, Vikas
APPLICANT: Li, Xuerong
TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
FILE REFERENCE: S12377019
CURRENT APPLICATION NUMBER: US1/10/087,464
PRIOR FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 06/272,930
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 59
SOFTWARE: PatentIn version 3.0
SEQ ID NO: 10
LENGTH: 1639
TYPE: PRT
ORGANISM: *Plasmodium falciparum*
US-10-87-464-10

Query Match 5.4%; Score 218; DB 14; Length 1639;
Best Local Similarity 18.6%; Pred. No. 0.0014; Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

Qy 23 KDTTG-VERHQQENESIK----KSSFTIDRNISTIRDFENPKOLKKLKKFREVDD 75
Db 232 KDNVGKMDYTKRNKNTENTNELLBESKTTDKNKNATKEE-----KKKLYQAOY 283
Qy 76 FISSETSGRMEEQYDQYDKGKNIAYDGTDEAYETEKLDEKSKYIGVLSPSKDGHFEL 135
Db 284 DLSIVYKQLES-----AHLILSV-----BR---KRDITK-----KNEKELL 319

Qy 136 GKLISVNSKNAKVYGNVYKSEIKAQTYDFH-----SKNTFDYANINDIVSGLA 188
Db 320 DKINEKQNPAPPANSQNTPTLDDKNNKLEERHEKEIKAQTKFNIDSFLFDPL----- 373

Qy 189 GDMRLFVTKDNQPKKAETKIRMEKEKETKSYMP-----YVSYGNV-----TELG-EGLSK 239

Db 374 ELEYVIREKQ-KNIDISAKTETKESTEPNPKYPNQVTPYPLSYNDINNALMELNSRGLI- 430

Qy 240 NPKDNNTKMEGSKYIDSEKQYL-LKDNTIL-----RKGYALKVTTYNPKTDML 289

Db 431 -NPFDTYKERSPKNITDNERKPKFINETKIEKIEKCKIESPKSYDRSKSLN----DT 485

Qy 290 EG-----NGVSK-----DIAKIQ-----NPNRALSETTIVADSRSRNEDGSTOSV 334

Db 486 KEYEKLNIEYDUSKFNNIDITNPFOMGKRYKSYKVEKLHNTFASYENSKHNLKLTK 545

Qy 335 IMAALDGF----NII-----RYQFPTKMKNGGEAIDKGDRGLVUTSSKLVFGKODKEYT 385

Db 546 ALKYMEDYSLRNIIVKEKLKYKLISKEMIEETE-NIKKUBQPLR-----ERKIT 598

Qy 386 GEDKFNVEAILEDGSMFLIDPKPVNLSMDKQYFNPSKSNSKTYWRPFPYLRGKIDKGGP 445

Db 599 IDENKDEKELIPLVSDIVKQVOKULL-----MVKIDELKKTQOLIKN-----LGHNTIHVPNSY 653

Qy 446 NWEILRNESVTDVNLKYGDLHDNTDFNQ-----NVR----- 480

Db 654 KOENKQK-----PYYLVLUKKEPKDQLKRVFMPKIVESLNEEKNKIKTEGQSPNSEPSTEI 709

Qy 481 -----DGDIMDWGMKDYK-----ANFRPDKV 501

Qy 710 TQGATTKPGQGQNSALEGDSYQDQEQKQAOQPVVUPVUPVUPVAKAQPTPPRPNKTEV 769

Db 502 TMDGSMVYLOGYSDINA-----KAVGVHQLFLYDVKPKENPDKGNTSEYADGKSW 556

Db 770 SKLD-----YLEKLYEFLNTSYCHKVILVSHSTNEKILKOYKITEEESKLLSCDPLLL 826

Qy 557 ENIN-----DRANGDG--BIOEAYIYKKEYTSFNDIKQDILKLNKIV 603

Db 827 FNIQNLNPVMSMFDNSNSQSQFLMUYEYEMVCHLYKNDKDKNLLPEAKVTSV 886

Qy 604 KDFARNITVKEFPLNKDGTGEVSELPKPHRVVTIONKEMMSSTIVSREDFPLPVYGELEK 663

Db 887 KILSSSMQFISLTPQDPKPEVSANDTSHSINNNSLKFNLIS-----LGRNK 937

RESULT 9-US-10-282-122A-70251

Sequence 70251, Application US1/0282122A
Publication No. US20040029129A1

GENERAL INFORMATION:

Qy 664 GYQFDGWEISQFEGK-----KQAGYVINLSKDTPKIVPKKIEKKEENKPTF-D 713
Db 938 IYQ---ELIGQKSSBNFVYKILKQDSTFVNESTPNPKVSKADDINSLDESKRKLLED 993

Qy 714 VSKKQNPQVHSQSLNEHSHRKEDQREEHQSOKSDSTKDVTATVLDKNNISK-STTNP 771
Db 994 INKLKKTQIQLSFDLYNKYKUKLERFDKKITVQYKQMOQKLTUKEQLESKNSLNNP 1052

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Hasebeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US1/0/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO: 70251

LENGTH: 1141

TYPE: PRT

ORGANISM: *Staphylococcus aureus*

US-10-282-122A-70251

Query Match 5.3%; Score 212; DB 15; Length 1141;
Best Local Similarity 20.9%; Pred. No. 0.0019; Matches 182; Conservative 116; Mismatches 336; Indels 238; Gaps 44;

Matches 182; Conservative 116; Mismatches 336; Indels 238; Gaps 44;

Qy 33 QENERSIKE-----KSSFTIDRNISTIRDFENPKDQLKUKKFR-----EVDDFTS 78

Db 221 KNPKEKELVNDNTDRSTRPKVATAPSVAQRLNKAQFAVAQPAVANNVLIT 280

Qy 79 ETGKMEBVLKYKQDKGKNIAYDGTPLVETE-----KLBKISKYVLS 125

Db 281 VTKOTK-----VGDCKDVAHAHDCKDIEYDFTIDKVKQGDIMTINYDKNVPISDLT 336

Qy 126 PSKD-----GHPELIGKISNVSKAK--VYIGNVYKIEKATVDFHSKTMFDL 174

Db 337 DKNDPIDIOTPSGEVIAKGTFDKATKQTYTFDVKYEDKARLJLYST-----IDK 390 ; PRIORITY APPLICATION NUMBER: 60/191,078
 Qy 175 YANINDIVGJAF-----GD--MRLFK-DNDQGKABKIRM- 209 ; PRIORITY FILING DATE: 2000-03-21
 Db 391 QAVPNETSLNLTATAGKETSONSVSYQDPNPKGNSNIOQIIFTKUDENKOTIEDQIYVN 450 ; PRIORITY APPLICATION NUMBER: 60/206,848
 Qy 210 PERIKETKSEY---PVSSYONVIEGEGD--LSKQKPDNLTKM-----ESGKYSD 256 ; PRIORITY APPLICATION NUMBER: 60/230,347
 Db 451 PLKKTATNTRKDIAGSQVDDQYK-ILGNGSTIIDQTEIKVYKVPNQQLPQSNRKYDF 509 ; PRIORITY APPLICATION NUMBER: 60/207,727
 Qy 257 SEKQOYILK-ONILRGKJYALKVTTNGKTDMLEGIVYKEDIAKIQKANPNLALSE 315 ; PRIORITY FILING DATE: 2000-05-26
 Db 510 SOYEDVTSQFDN---KKSFSNNVATLDFG---DINSAYI-----IKVSK 548 ; PRIORITY APPLICATION NUMBER: 60/230,347
 Qy 316 TTYAISR-NVEDGRSQTOSVMSALOGNIRYQVFKMDKG----- 358 ; PRIORITY FILING DATE: 2000-09-09
 Db 549 YTPTSDEBLDAQTSKRTT--DKYQTYNYGCSNFTVTSIDTGGDGTVPBPEKLYIG 606 ; PRIORITY APPLICATION NUMBER: 60/242,578
 Qy 359 ---EATDKDGSNLVTDSSKLVJLPGKODKEYGEDKENVATEKEDGMSLFDITKPVNLSD 414 ; PRIORITY FILING DATE: 2000-11-27
 Db 607 DVMWEDVTDKG-----VQGTSKESKPMNL-VTLYTPDT-----TKSVRTDN 650 ; PRIORITY APPLICATION NUMBER: 60/257,931
 Qy 415 KNY--ENPSKSKNKLKVY--RNTPYRKLKIS-----DKGFPNWLARVNESSVNDNTLYG 463 ; PRIORITY FILING DATE: 2000-12-22
 Db 651 GHIEFGGLKDGTITVKEPAPGALPFTKVNGETDGEKDSNSSITVKNG-----D 702 ; PRIORITY APPLICATION NUMBER: 60/267,636
 Qy 464 DHHTD-----NTRDFNITKLNKVKDGG--DIMWGMKDYGKANGFPDVKUTMDGNTYLT 512 ; PRIORITY FILING DATE: 2001-02-09
 Db 703 DMSLDTGPKYKPKYNGDQYVNDTNGKQDNEPGRKVT-----LKSTGKV-IGT 756 ; PRIORITY APPLICATION NUMBER: 60/269,308
 Qy 513 GYSDLMAKAVGVMYQFL-YDNVYKPEVND-PKG-----NTSIEYAD--GGSVFNINDK 562 ; PRIORITY FILING DATE: 2001-02-16
 Db 757 TTIDASGK---YKFPLDNGNTYTFETPAGYPTPWTNTAEDKNSGJLTGVIKA 811 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 Qy 563 RNNGFGEBIQBOHYY-----NGKEYTSEFDIKQDILKTANIK-IVVKDFAANT 610 ; NUMBER OF SEQ ID NOS: 78614
 Db 812 DMTLTDGPKYKPKYSLGDYVWYDSNKQDGQDSTEKGJIKDVKVTLNEKGIVGIFTKDE 871 ; SOFTWARE: Patentin version 3.1
 Qy 611 TWRKFILNKDGEVSEL--KPRHVTUON-----GKEMSTSIVSERDIPLYVKG 659 ; SEQ ID NO: 76365
 Db 872 NCKYRFDNLDSSCKYKVFPEKPKAGLDTOTVNTTDDKODADGKSEVDTTDDHDFI----- 925 ; LENGTH: 4688
 Qy 660 EBLKGXOPDGMELISGFECKKODAGYVNLSKDTPIKPVKKLBKKEBENKTPDVSKKKD 719 ; TYPE: PRT
 Db 926 -LNGYTFEEDPSDSDSDSDSDSDSDSD-----DSDSDSDSDSDSDSD 975 ; ORGANISM: Ureaplasma urealyticum
 Qy 720 NPOVNHSQLNESHRKEDLQREBHSQKSDSTD 751 ; US-10-282-122A-76865
 Db 976 SPSDSDSDSDSDSDSDSDSDSDS 1007 ;

RESULT 10

US-10-282-122A-76865
 Sequence 76865, Application US/10282122A
 Publication No. US20040029129A1
 GENERAL INFORMATION:

APPLICANT: Wang, Liangbu
 APPLICANT: Zamudio, Carlos
 APPLICANT: Malone, Charyl
 APPLICANT: Ohlsen, Kari
 APPLICANT: Zyskind, Judith
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John
 APPLICANT: Carr, Grant
 APPLICANT: Yamamoto, Robert
 APPLICANT: Forsyth, R.
 APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELTRA 034A

CURRENT APPLICATION NUMBER: US/10/282.122A

CURRENT FILING DATE: 2003-02-20

Db 226 YGNVIELEGEGDLSKNUKDNLTKMESSKLYSEKQOYLKONILRGKJYALKVTTNGK 285 ;
 Db 291 KNNV-----GYENTKPLNPKYELMIVNNELNDVLPNELME-HKVSUNKPTK 344 ;
 Qy 286 TDMLEGNGVYKSEKDLAKIOKAMP--NRLAUSSETTYADSRSNVEDRGRSTOSVMSALDF 342 ;
 Db 345 VNFIDNKVY-EVDLGVQLKNTPLETLEDLHQYKINKAKTDEKGRAVFDI--SSLGDN 401 ;
 Qy 343 NTRYQFTEKNDKGKHAID-----KOGNL-----VTDSSKLUFLGKDK 383 ;
 Db 402 NLI--YEVIGKRNENVWVNQKIPYHRNTRSTNLNMSALNTPYQYKNGDNLIAKAPY 459 ;
 Qy 384 YTGEDKENV-----BAIKEDGMSLFDITKPVNLSDKMYFFNPKSKNKAIVNP 431 ;
 Db 460 YNQOQVGFIDKDNQEHQOLAKVKGDTAF-DPGALN----NNSNYSLDKTVVSNSP 513 ;
 Qy 432 E-----FYLRK--1SDKGFPNWL-----RVAESVVDYLI--YGDJLHDNT 470 ;
 Db 514 QNVLVSNFDLSKQKOLIKKEANASVSDSTKQTQILANLNDLNLINQMLVATPD--NND 570 ;
 Qy 471 RDNFTKLNKVKDGD--DIMWGMKDYGKANGFPD-----KYMDDGMYVYLOGYSDNANKA 521 ;
 Db 571 KEYKVVANVDQNNKIEFD-----SNDLPGKJYTHLAKCENNDLNKVKINLDFELKDKI 623 ;

Query Match 5.1%; Score 207; DB 15; Length 1191;
 Best Local Similarity 19.6%; Pred. No. 0.0038;
 Matches 178; Conservative 139; Mismatches 275; Index 314; Gaps 43;
 Qy 1 KUGETAESKEFLNGKKEG-.....-SILKDDTGVERHHQENHESTI 39
 Db 120 RJKD1QB-LFMDTGIGKEGYSTIGQSKTEAVLISGKPEERALLBEAGIVKFRTRVDA- 177
 Qy 40 KEKSSFTIDRANISTIRD-.....-FENKDJKKLKIKKFREVD- PTSEIGK- RMEE 86
 Db 178 -EKKLLENTNOMLVRINDLIRTYEEERLPLRTESEKARFVLESDELKTCETKEMTIYSDN 236
 Qy 87 YDYKUDKGKNTIA-.....-YDDGT-.....-DLEYETEKDGI-.....-KSIKYGLSP 126
 Db 237 IDYRINDLQKOMADLKLISIDENVKDKEKISLKVATESLUDFDAYKSSNKTY- E 292
 Qy 127 SKDGHEFILGI-.....-SNVSKQAKVYGNNSIE-.....-IKATYDPEHSTKMTFELY 175
 Db 293 SKSEHOKTILSHIELLEKKEKTSNSDVAKNLYKREIDLDNSTVNL- SREYEQLKTLD- 349
 Qy 176 ANINDTVDGLAFAGDMRLFVQNDQKKAIBKIRMP-.....-KIKETKSE-.....-YPVSSGN 228
 Db 350 KNVNK-.....-EULSKNNKSEEKKNONGLIESWEKSTQKQYKODAIDITISQNN 400
 Qy 229 VIELGBGDLSKNPKDNLTKMES-.....-GKIVDSEKOOYLUUDNTILRGYALKVUTNPG 284
 Db 401 EVVI-.....-LKKIESNESKESKIKRAGEGYSKS-.....-LKNENVKTLSBELVKLN- 448
 Qy 285 KTDMLLEGNGVYSKEDIAKIQCANPNRALSSTTIVADSRNEDGRSQTQSVMSALDFNI 344
 Db 449 -DKLISGYENQIRENRSKISKLN-.....-RITSD-.....- 475
 Qy 345 IRYQVFTKMDKGEAIDKODNLVTDSSKLYLFGKUDKEYTGEDKNEVAKEDGSMFLI 404
 Db 476 -.....-EKUNRELANSKSNKLEANQMLNL-.....-EKQVGGYNR-SVKNLQHVTGKFV 522
 Qy 405 DTKPVNLSDMDPNTENPSKSNSKLYVNRNPBFYRKGTSDDKGHHFWELRVNESVVD- NYLI 461
 Db 523 DVKPPESSFVLGKEVKVKKEFETAV-.....-EISLGAISD-.....-LITLDDNIAKLLNLYK 573
 Qy 462 YGDIHDHNTDTPFI-.....-KLNVKDG-.....- 482
 Db 574 SKNLRGATPLPLNLTIKGRKONISADATHERKFRIGIASELIBKDSLTELPAVNTVLGRTV 633
 Qy 483 DIMWGMKDYANGFPDKVUTMDGNYLTQGYSLDNAKAVGVMYQFLYDNKPENVNDPK 542
 Db 634 DMDMSALKIAKLNSTSFKIVTILTGEV-.....- 663
 Qy 543 GNTS-.....-IEYADGKSUVFNINDKRNNGPFGELQEOHITYINGKEYTSENDIKOIDKTLIK 600
 Db 664 GSITGGSTYSKASII-.....-GRKRETEELNLBNNSQQA-.....-LEOSSNKLLENK 710
 Qy 601 IWKDOPARNTVKEFLINKOTGEVSLKPHRVTIUTQNGKEMSSTTVEEDRFLPVKGE 660
 Db 711 KWKVKE-.....-LDNLCLD-.....-LTDTT-HGEKIELTKIKERIKMSIDLESEK 750
 Qy 661 DEKGYQFDGWHWISGFEGKKDAGYVINVLSKDTPIKPVFKTBK-.....-KEEENKPTFDYSSK 717
 Db 751 IANKSTNTSVEIG-.....-FIKEKINKHBLKLYKKEENKA-.....-LKLK 789
 Qy 718 KDNPQVNHQSOLNESHRK-.....-EDLQREHHSQSD-.....-STKVDTAVLDK 759
 Db 790 EAN-.....-NNNLIDELERLKLKDENSKVNLNBHEIMSMKVDAKSDDEMMSSTRBIERYKEM 846
 Qy 760 NNISSK 765
 Db 847 HNMENK 852
 RESULT 11
 US-10-282-122A-52048
 Sequence 52048, Application US/10282122A
 Publication No. US20040029129A1
 GENERAL INFORMATION:
 APPLICANT: Wang, Liangsu
 APPLICANT: Zamudio, Carlos
 APPLICANT: Malone, Cheryl
 APPLICANT: Hasebeck, Robert
 APPLICANT: Ohlsen, Kari
 APPLICANT: Zyskind, Judith
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John
 APPLICANT: Carr, Grant
 APPLICANT: Yamamoto, Robert
 APPLICANT: Forsyth, R.
 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 FILE REFERENCE: EULTRA.034A
 CURRENT APPLICATION NUMBER: US/10/282,122A
 CURRENT FILING DATE: 2003-02-20
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/230,335
 PRIOR FILING DATE: 2000-09-06
 PRIOR APPLICATION NUMBER: 60/230,347
 PRIOR FILING DATE: 2000-09-09
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/267,636
 PRIOR FILING DATE: 2001-02-09
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 Remaining Prior Application data removed - See File Wrapper or PALM.
 NUMBER OF SEQ ID NOS: 78614
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 52048
 LENGTH: 1191
 TYPE: PRT
 ORGANISM: Clostridium acetobutylicum
 RESULT 12
 US-10-741-616-7

Sequence 7, Application US/10744616
 Publication No. US20050026170A1
 GENERAL INFORMATION:
 APPLICANT: Parti, Joseph M.
 APPLICANT: Foster, Timothy J.
 APPLICANT: Hock, Magnus A.O.
 APPLICANT: Eidhian, Beirdre N.
 APPLICANT: Perkins, Samuel L.
 TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from *Staphylococcus aureus*
 FILE REFERENCE: P06281US2/BAS
 CURRENT APPLICATION NUMBER: US/10/744,616
 CURRENT FILING DATE: 2003-12-24
 PRIORITY APPLICATION NUMBER: 60/066,815
 PRIORITY FILING DATE: 1997-11-26
 PRIORITY APPLICATION NUMBER: 60/098,427
 PRIORITY FILING DATE: 1999-08-31
 NUMBER OF SEQ ID NOS: 23
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO: 7
 LENGTH: 1166
 TYPE: PRT
 ORGANISM: *Staphylococcus aureus*
 US-10-744-616-7

Query Match 5.0*; Score 203; DB 17; Length 1166;
 Best Local Similarity 21.6*; Pred. No. 0.062; Gaps 53;
 Matches 184; Conservative 120; Mismatches 291; Indels 258; Gaps 53;
 Qy 33 QNEBESIKE-KSSFTIDRNISTIRDRENKDKLKKKKR-----EVDDFTS 78
 Db 226 KNPKEKUKELVANDSNTHDSTKREVATAPTSVAKRWNKMRPAAVQPAAVSNVNDIK 285
 Qy 79 ETGKRMEEYDVKYDKGNIIAYDGTDELEYET-----KLEDEKSKIVGVL 125
 Db 286 VTKQTIK---VGDKONVALAHDQDIBYDTEFTIDNKVKKGDWTINYDKNVIFSDLT 341
 Qy 126 PSKD-----GHEFELIGKSNVSKNAK---VYVGNYKSYEIKATKQYDFHSK---- 168
 Db 342 DKNDPDPIDTDPSEGVIAKGTDFKATQTYTFDYDVKYEDIKSRMLYSYTDKKTVPNE 401
 Qy 169 ---TMTF---DLYANIN-DVGLAPAGD--MRLFVK-DNDQGAKIWRM-PFK 214
 Db 402 TSILNLTTATAGKETQSNTVQDPMW-GDSNITQSTFKLJEDQKTEQDQIYVNPILKS 460
 Qy 215 ETKSEY---PVYSSYGNVILEGEGD--LSKQKPDNLTKM-----ESGKYSRSBHQ 261
 Db 461 ATWTKVTDIAGSQVDDYGN-I-KUGNGSTIDQNTBIVKIVKNSDQQLRQPSNKRIDFSQED 519
 Qy 520 VTSQFDN---KGSFSNNVATLDFG-----DINSAYI-----IKVWSKYPTS 558
 Db 321 DSR-NVEDGRSTOSVULMSALDGENIIRYQVFTPKMDKG-----E 359
 Qy 262 VILK-DMLTRKGYALKUTTYPGKTMLEMGWVSKEDIAKLQKAMPNTRALESETIYA 320
 Db 459 DEDBLDIAGTSRHTT--DKYGVVNYAGVSNFTWTSNTGGDGTVKRERKJYKIGDVME 616
 Qy 360 A1DKDGNLVLTDSSKLVJFGKDKKEYTSEDFKENVTEAKEDGSMFLIDTRKPNVMSMDKRY-F 418
 Db 617 DVDKDG-----VOGTSKEKPMANL_VLTFPDGT---TKSVRTDANGHIEP 660
 Qy 419 NPSKSNKLYV---RNPRFYLGRKIS-----DKGGTENWELARVNESTVNDNLYIGDHAID 468
 Db 661 G3LKDGETTYTKEFPTGFLPTKINGTIDGEKUSNGSSVTVKINGK-----DMSLD 712
 Qy 469 -----NTEDFNKLANKVKG--DIMDWGMKDYKANGFPIKUTDMGNNVLTQGYSDL 517
 Db 713 TGPYKEPKYNTQDLYWEDTNKOGIQDANEPIGKIDVKY-----LKOSTGKJ-IGTTTDA 766
 Qy 518 NAKAVGHWYQFLJDNKPEVNIDPKGNISIEV---ADGKSWVN--INDKRNQNGFGEIQ 572
 Db 767 SGK-----YKF-----TLD--NGVYTERETPACVTPYKNTADDKOSNGL----- 807
 Qy 573 EOHIVYINGKEYTISFNDIKOIQIDKTNLNIKIVKDFARNI---TVKEFI---LNKDTGEVSEL 627

RESULT 13
 Sequence 7646, Application US/10032585
 Publication No. US20030180953A1
 GENERAL INFORMATION:
 APPLICANT: Terry, Roemer D.
 APPLICANT: Bo, Jiang
 APPLICANT: Charles, Boone
 APPLICANT: Howard, Burrey
 TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
 FILE REFERENCE: 10182-005-999
 CURRENT APPLICATION NUMBER: US/10/032,585
 CURRENT FILING DATE: 2001-12-20
 NUMBER OF SEQ ID NOS: 8000
 SOFTWARE: Patentin version 3.1
 SEQ ID NO: 7646
 LENGTH: 1881
 TYPE: PRT
 ORGANISM: *Candida albicans*
 FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (1881)..(1881)
 OTHER INFORMATION: X-any amino acid
 US-10-032-585-7646

Query Match 5.0*; Score 203; DB 14; Length 1881;
 Best Local Similarity 20.8*; Pred. No. 0.011; Gaps 39;
 Matches 172; Conservative 132; Mismatches 300; Indels 222; Gaps 39;

Qy 22 KDTTGYE---HHHQNEESEIKESKSTORN-ISTIRDRENKDKLKKTFREVD 76
 Db 1140 KTDATDEETKTKHITDQBERAKSOSKSLXHJDTEDEKREKIKLETTKENEETMFEK 1198
 Db 1140 KTDATDEETKTKHITDQBERAKSOSKSLXHJDTEDEKREKIKLETTKENEETMFEK 1198
 Qy 77 TSETGKMEETYKIDKGNIIAY--DDGTDELEYETKDETKSKYIVGVLSPSKDGHF 134
 Db 1199 KTELANNKLTQBEKISOLRTSVAISDKSKSLXHJDTEDEKREKIKLETTKENEETMFEK 1258
 Db 1259 KEQLOVNDKCK-----ELEAC-----LKKLTETKEKEINDLIRKLEA----- 1297
 Qy 195 VTKNDPKKAETKIRMPPEKIKETKSEPYVSSYGNVILEGEGDLSKONPKPDNLTKMEEGKLY 254
 Db 1298 KSDHDTREKKSL-----LIEDPKSE-----SEBNVNLNE-----QBEKLK 1334
 Qy 255 SDSEKOQULLKONLILRGYALKVTTYPGKTM-----LEGNGVSKEDIAKIQLK 305
 Db 1335 GEREKE---VRD---TQSOLAKTTDWEKIKTDLKULKEKSDEKLT---NKESVDTLKK 1385
 Qy 306 ANPFLR---ALSETTIVADSRNVEDGSTOSVULMSALDGENI-----IRQVQF- 350
 Db 1386 EVENLKGEISLEQDKDITKVKELAQPLTKTSNLDSTIMELEKTELEKKVNLTE 1445
 Qy 351 ---TEFQNDKG---EAIKDGNLYTDSKSLVLRGKDKKEYTGDEKPKVNAIKEDGSM 402
 Db 1446 ATSELTLQDQNQSLTEIEKTKAALKKSS-----DLEVGNOKEQ-----DSLK 1493

QY 403 RDTKPKNLSDKDNVNPNSKAKIYVNPFCY---LRGKISDKGKGFNPMLRVNWSVDP 457
 Db 1494 SVKSELENKENPKENQETTSLSKAKIEEKQETKIVTQLOTELKORISEVERERAMLESENSETV- 1552
 QY 458 NYLIGDHLHINTRDENKIKUNVKDGDIMDWSMKDVKANGFEDPKVTDGMVNLQGDSL 517
 Db 1553 -----IKEYS---DKISLE---SKI 1567
 QY 518 MAKAVGUVHYQFLDNYKPEVNPDKPGNTSISY---ADGSKVVMNINDRNGNDFGEI 571
 Db 1568 NS-----IKNHISKEITTHNEOKTSLSKDIKLQSQDRESAQTOLEDGENQNLKKA 1618
 QY 572 Q-EQHTYINGKBYTSF---NDIKQ1ID---KYNLNIKUVKOFARNTVKEB---ILNKO 620
 Db 1619 SLEKH---NTSATSIEEKKNQIKELSETIMSLKBLTQSALKOSQKEYTKLTKNSD 1675
 QY 621 T---GEVSLKPKHPTVTTQNGKEMSTIVSEEPILPVGLEKGQFQDFGWEISG 675
 Db 1676 TESKLEROLELE---KVSKD1QTADEKUGJTERETAL---KSELET----- 1717
 QY 676 ESGKDKDAGVVLKSKDTPKPKKIEERKESENKPPDVSCKDNQVNHSQLNESHRK- 734
 Db 1718 --VKNGLISTTSELAALTKTW---KSLERKEKLQFELQGNSKTELEDYIQKHSIDISKAL 1774
 QY 735 EDLQREHNSQSDSTK-----DVTAVLQDKNNSKSTTNPNK 773
 Db 1775 TDELKEKTKQDDSKKKLTELJENDLST---KKELETEKOTSFK 1817

RESULT 14
 US-03-742-096-3
 ; Sequence 3, Application US/09742096
 ; Patent No. US20020155441A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DRUILHE, PIERRE
 ; APPLICANT: DAUBERIES, PIERRE
 ; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
 ; FILE REFERENCE: 200773US04050IV
 ; CURRENT APPLICATION NUMBER: US/03-742, 096
 ; CURRENT FILING DATE: 2000-12-22
 ; CURRENT APPLICATION NUMBER: US 08/973, 642
 ; PRIOR FILING DATE: 1998-02-06
 ; PRIOR APPLICATION NUMBER: PCT/FR96/00894
 ; PRIOR FILING DATE: 1996-06-12
 ; PRIOR APPLICATION NUMBER: FR 95/07007
 ; PRIOR FILING DATE: 1995-06-13
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 3
 ; LENGTH: 1786
 ; TYPE: PRT
 ; ORGANISM: *P. falciparum*
 ; US-03-742-096-3

Query Match 5.0%; Score 200.5; DB 9; Length 1786;
 Best Local Similarity 20.5%; Pred. No. 0.015; Gaps 52;
 Matches 204; Conservative 160; Mismatches 329; Index 301; Gaps 52;

QY 6 AES-KPKN-LGNGKKGSKLKD-----TGVERHHOENE-----ESTK 40
 Db 868 AESVTVPNSNLLBEEQENTITDIEKBEHENVSALENTOSEBKEVIVDEEK 927
 QY 41 EKSSFTI-----DRNISTRD-FEN-----KDLKLUKKPFREYDFTS 78
 Db 928 BEVATLIELTWEQAEEKSANTITEIPEENLENAEVENENVLENKLENETVNLKVE 987
 QY 79 ET---GKRMG---EYVKY---DDKG---NLYDDGSDLEET-----EKL--- 114
 Db 988 EWEVLSGESLHNENMKAFFSBEIDNVKGJGQBNLTT---GMRSTTSIVQSEEKVDLN 1044
 QY 115 -EIKSKYGV-SPSKQKHFETLGKLSNVSKWAV-----YYGNVKSIBIKATKDH 166
 Db 1045 ENWVSLDNTENMKG---LANKLENISSEBSQVOTVTEVNEQNTVWDVDPANK--- 1097

RESULT 15
 US-10-415-253-2
 ; Sequence 2, Application US/10415253
 ; Publication No. US20040067236A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cohen, Joe
 ; APPLICANT: Drulhe, Pierre
 ; TITLE OF INVENTION: Immunogenic Compositions Comprising
 ; TITLE OF INVENTION: Liver Stage Malarial Antigens
 ; FILE REFERENCE: B42520
 ; CURRENT APPLICATION NUMBER: US/10/415, 253
 ; CURRENT FILING DATE: 2003-04-25
 ; PRIOR APPLICATION NUMBER: PCT/EP01/12349
 ; PRIOR FILING DATE: 2001-10-23
 ; PRIOR APPLICATION NUMBER: EP00203724.0
 ; PRIOR FILING DATE: 2000-10-25
 ; SOFTWARE: Fast-SEQ for Windows Version 4.0
 ; SEQ ID NO 2

QY 167 SKMTFUDLYANINDIVGLAFAGDMRLFVKNDOKKAKIRKIRMPKKEKTYSEPVYSSY 226
 Db 1098 -----DOFGLGILNEBAGL-----KEMPFENLEDVURKESPSDVITVERIKDQPVQKEKE 1146
 QY 227 GNVIELEGD---LSKKNKPDPNLTKM---ESGKYSDSERKQQVLLKDNILRKGYALK 277
 Db 1147 VSIEEMMENIVDVLBEEKSDLTDMIDAVESEISISDSKEETESIKDK---EKDSLV 1203
 QY 278 VTTYNPG-----KTDMLEGNGVSK---EDIARIQKANPL----- 310
 Db 1204 VEEQDNMDMSVEVKULKOMBEILMDAVEINDTSKJETEOLNEFADLKDMEK 1263
 QY 311 -----RALSETTIYADSSRNVEDGR---STQSVLMSALDGEMNIRYQVFTKNDKGEADK 363
 Db 1264 IKEYLKLSE---DSKEI1DAKDTLEKVEERD-----ITTLDEVELVLDV 1309
 QY 364 DGNLVTDSSKVLFLGKDKKEYGEDKPNVNEAKEDGSMFLRDTKPNL-----SM 413
 Db 1310 EEDKLEKVSDL---KDLDEEDILKE---VKEIKELESELEDEYKELKTTEDILEBKEI 1362
 QY 414 DKYNEFNSKAKIYVNPFCYVPEFLRGKTSKDGCGFNPWNLRSVWV-----NYLIVGDL 465
 Db 1363 EKDHFEKPEERABEIKOLEADILKEV-----SLNEBEKKKEFHELKSEVEHITGDA 1417
 Db 1418 HIKGLBEBBLE-BVDDLGSTIDMLKGDMDKESLDVTTKLERV---ESIKDVL 1473
 QY 519 AKAVGUVHYQFLYDN-----KPEVNIDPKGNTSIEYADGSKVFNINBK-RN 564
 Db 1474 SSALGMDEBQMKTRKAQRPKLEEVILKEMVKEPKKUT-----KKVQFDFIKOKEPK 1528
 QY 565 NGFDGETIQECHIYINGK-----EVTSFNDIKQ1DTKTN-----IKTVKDFAR-N 609
 Db 1529 EIVEVZMKDDEBDEVEDDEBDEEDBEDKVEDIDEDEBDEIDEDGKDVBUDIIVQKREK 1588
 QY 610 TIVKEFILINKDGTG-VESELKPH-----RVTUTONGKEMSTIVSEED 651
 Db 1589 VKAKKKKLEKKEVVGSKLKKRHDVEMVKYQOKIDKEVDEKVSKALESKNDVNTVKQNQD 1648
 QY 652 FILPVYKGEBLKQYQFDGMWISGFGFQGKQDAGYVIM-----SKDTFKPKVFK 698
 Db 1649 FFSKV-KNFVCKYVFAFPFSAVA---FASYVGFRTFLFSSCVTILASSTYLSKVD 1705
 QY 699 KTEKEKEENK-----TDFV-----SKKKNPQVNHSQLNESHRKDQLR 739
 Db 1706 TINKKE-----RPFYTFVDFDIFKLNKLYLQQMKEKSKEKNNVIRT- NKAEGKGNVQ 1760
 QY 740 BEHSKDSDSTKDVTAVLQDKNNSKSTTNPNK 773
 Db 1761 TNKTEK-----TKVDKNNKVKPKKRTQSK 1786

; LENGTH: 1787
; TYPE: PRT
; ORGANISM: K1 Parasite Clone
US-10-4152-253-2

This Page Blank (uspto)

Scoring table: BLOSUM62										Alignments									
Searched: 1612378 seqs, 512079187 residues										Database: Gapox 10.0 , Gapext 0.5									
Total number of hits satisfying chosen parameters: 1612378										Post-processing: Minimum Match 0% Maximum Match 100% Listing first 45 summaries									
Database : UniProt 03::*										1: uniprot_sprot:*									
2: uniprot_trembl:*										Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.									
SUMMARIES										RESULT 1									
Result No.	Score	Query Match	Length	DB	ID	Description				Q9AHTS	PRELIMINARY; PRT; 2119 AA.				RT	DT	DT	DT	DT
1	4026	100.0	2119	2	Q9AHTS	Q9AHTS; 01-JUN-2001 (TrEMBLrel. 17, Created)				Q9AHTS;	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)				DR	DR	DR	DR	DR
2	4026	100.0	2140	2	Q9TRX6	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)				Q9TRX6	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)				DE	Serine protease (fragment).			
3	3793.5	94.2	2144	2	Q9SAM8	Name=ptcA;				Q9SAM8	Name=ptcA;				GN	Streptococcus pneumoniae.			
4	3709.5	92.1	2144	2	Q8DQPT	OS				Q8DQPT	OS				OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
5	270.5	6.7	2649	2	Q7RAS7	OC				Q7RAS7	OC				NCBI_TaxID=1313;	NCBI_TaxID=1313;			
6	265.5	6.6	1642	2	Q8IB84	RN				Q8IB84	RN				SEQUENCE FROM N.A.				[1]
7	264	6.6	1504	2	Q8IL45	RX				Q8IL45	RX				STRAIN=N4; MEDLINE=21116976; PubMed=11179332;				RX
8	259	6.5	1850	2	Q7RG92	RA				Q7RG92	RA				DOI=10.1128/IAI.69.3.1591-1598.2001;				RA
9	259	6.4	2157	2	Q7RR99	RA				Q7RR99	RA				Wizemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C., Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomonen E., Gayle A., Brewan Y.A., Walsh W., Barron P., Lathigra R., Hanson M., Langemann S., Johnson S., Koenig S.;				RA
10	258.5	6.4	1811	2	Q7RBB9	RA				Q7RBB9	RA				"Use of a whole genome approach to identify vaccine molecules affording protection against Streptococcus pneumoniae infection."				RT
11	257.5	6.4	2661	2	Q7RMS4	CC				Q7RMS4	CC				Infect. Immun. 69:1593-1598 (2001).				RL
12	257	6.4	1389	2	Q7RBU4	DR				Q7RBU4	DR				-1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (BY similarity).				CC
13	255	6.3	1777	2	Q8IP34	DR				Q8IP34	DR				EMBL: AF291699; AACU9159.1; --.				DR
14	253.5	6.3	1127	2	Q9YV66	DR				Q9YV66	DR				MEROPS: S08.064; --.				DR
15	252.5	6.3	2227	2	Q8I121	DR				Q8I121	DR				Q9YV66				DR
16	252.5	6.3	2723	2	Q7RQB6	DR				Q7RQB6	DR				Q9YV66				DR
17	251.5	6.2	2894	2	Q8IHY4	DR				Q8IHY4	DR				Q9YV66				DR
18	251.5	6.2	3381	2	Q8I2V4	DR				Q8I2V4	DR				Q9YV66				DR
19	251.5	6.2	3319	2	Q8IE65	DR				Q8IE65	DR				Q9YV66				DR
20	250	6.2	1474	2	Q8ILU2	DR				Q8ILU2	DR				Q9YV66				DR
21	250	6.2	1850	2	Q7T3Z8	DR				Q7T3Z8	DR				Q9YV66				DR
22	247.5	6.1	5767	2	Q8IS25	DR				Q8IS25	DR				Q9YV66				DR
23	246.5	6.1	2033	2	Q8IM18	DR				Q8IM18	DR				Q9YV66				DR
24	245	6.1	2269	2	Q8IJ42	DR				Q8IJ42	DR				Q9YV66				DR
25	242.5	6.0	3264	2	Q7RBL0	DR				Q7RBL0	DR				TIGRFAM: TIGR01167; LXPX anchor; 1.				DR
26	242.5	6.0	3317	2	Q8EP98	DR				Q8EP98	DR				PROSIE; PS05047; GRAM POS_ANCHORING; 1.				DR
27	241.5	6.0	1033	2	Q8IBB8	DR				Q8IBB8	DR				PROSIE; PS05047; SUBTILASE_HIS; UNKNOWN 1.				DR
28	241.5	6.0	2694	2	Q7RUJ1	DR				Q7RUJ1	DR				PROSIE; PS05047; WD_REPEATS 1; UNKNOWN 1.				DR
29	241.5	6.0	3063	2	Q61MC1	DR				Q61MC1	DR				Q9YV66				DR
30	240.5	6.0	2740	2	Q7RFS2	DR				Q7RFS2	DR				Q9YV66				DR
31	240.5	6.0	4433	2	Q8IJ55	DR				Q8IJ55	DR				Q9YV66				DR
32	240	6.0	1455	2	Q8IKG8	DR				Q8IKG8	DR				Q9YV66				DR
33	240	6.0	10061	2	Q8I3Z1	DR				Q8I3Z1	DR				Q9YV66				DR
34	239.5	5.9	2569	2	Q8IBG8	DR				Q8IBG8	DR				Q9YV66				DR
35	239	5.9	2586	2	Q7PDT7	DR				Q7PDT7	DR				Q9YV66				DR
36	238	5.9	3322	2	Q8IXL2	DR				Q8IXL2	DR				Q9YV66				DR
37	237	5.9	238	2	Q868Y0	DR				Q868Y0	DR				Q9YV66				DR
38	237	5.9	3704	2	Q8IKY8	DR				Q8IKY8	DR				Q9YV66				DR
39	237	5.9	238	2	Q7RSQ8	DR				Q7RSQ8	DR				Q9YV66				DR
40	237.5	5.9	1104	2	Q8IC29	DR				Q8IC29	DR				Q9YV66				DR
41	237.5	5.9	3535	2	Q8IBI4	DR				Q8IBI4	DR				Q9YV66				DR
42	237	5.9	2511	2	Q7RTF4	DR				Q7RTF4	DR				Q9YV66				DR
43	237	5.9	237	2	Q8T242	DR				Q8T242	DR				Q9YV66				DR
44	236.5	5.9	1114	2	Q8IAL5	DR				Q8IAL5	DR				Q9YV66				DR
45	236.5	5.9	1738	2	Q8IAL5	DR				Q8IAL5	DR				Q9YV66				DR

SQ	SEQUENCE	2119 AA;	238227 MW;	51.7F9BYF6B950A6A	CRC64;
Query	Match	100.0%;	Score 4026;	DB 2;	Length 2119;
Best	Local	Similarity	100.0%;	Pred. No. 2e-161;	Indels 0;
Matches	773;	Conservative	0;	Mismatches 0;	Gaps 0;
ID	Q97RY6	PRELIMINARY;	PRT;	2140 AA.	
AC	Q97RY6				
DT	01-OCT-2001	(TREMBLrel. 18, Created)			
DT	01-OCT-2001	(TREMBLrel. 18, Last sequence update)			
DT	01-MAR-2004	(TREMBLrel. 26, Last annotation update)			
DE	Serine protease, subtilase family.				
GN	OrderadLocusNames=SPP641;				
OS	Streptococcus pneumoniae.				
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
OC	Streptococcus				
OX	NCBI_TaxID=1313;				
RN	[1]				
RESULT 2					
Q97RY6					
ID	Q97RY6	PRELIMINARY;	PRT;	2140 AA.	
AC	Q97RY6				
DT	01-OCT-2001	(TREMBLrel. 18, Created)			
DT	01-OCT-2001	(TREMBLrel. 18, Last sequence update)			
DT	01-MAR-2004	(TREMBLrel. 26, Last annotation update)			
DE	Serine protease, subtilase family.				
GN	OrderadLocusNames=SPP641;				
OS	Streptococcus pneumoniae.				
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
OC	Streptococcus				
OX	NCBI_TaxID=1313;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ATCC BA-334 / TIGR4;				
RX	MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;				
RA	Tettelein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,				
RA	Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,				
RA	Durkin A.S., Gilmartin M.L., Kolonay J.F., Nelson W.C., Peterkin J.D.,				
RA	Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,				
RA	Holtapple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,				
RA	McDonald L.A., Feldblyum T.V., Anguoli S.V., Dickinson T.,				
RA	Hickey E.K., Holt T.E., Loftus B.J., Yang P., Smith H.O., Venter J.C.,				
RA	Dougherty B.A., Morrison D.A., Hollingshead S.K., Frazer C.M.;				
RA	"Compete genome sequence of a virulent isolate of Streptococcus pneumoniae.,"				
RJ	Science 293:498-506 (2001).				
RJ	-1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by				
CC	an amide bond (By similarity).				
CC	EMBL: AE007373; AAC74791.1; -.				
DR	PIR; P5074; P55074.				
DR	HSSP; P0082; ZSBT.				
DR	MEROPS; S08.064; -.				
DR	TIGR; SP041; -.				
DR	GO; GO:0009986; C:cell surface; IEA.				
DR	GO; GO:0005618; C:cell wall; IEA.				
DR	GO; GO:008233; F:peptidase activity; IEA.				
DR	GO; GO:0004289; F:subtilase activity; IEA.				
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.				
DR	InterPro; IPR01045; DUF0134.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR003137; PA.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.				
DR	InterPro; IPR01259; Prot_inh_S8A.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF0280; DUF1034; 1.				
DR	Pfam; PF00746; Gram_pos_anchor.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR000209; Pept_S8_S53.	</			

QY	361	IKDGNIVTDSKLVIFGKDKDKEYTGEDGSMFLIDTPVNLSMDKVFNP	420	DR PRINTS; PRO0723; SUBTILLISIN.
Db	1694	IKDGNIVTDSKLVIFGKDKDKEYTGEDGSMFLIDTPVNLSMDKVFNP	1753	DR TIGRMS; TIGR01167; LPXTC anchor; 1.
QY	421	SKSNKTYVRNPFYLRKISDKGGFWELRVNESVVDNYLYGDLHIDNTRDFNKLNVK	480	DR PROSITE; PS00137; SUBTILASE HIS; UNKNOWN 1.
Db	1754	SKSNKTYVRNPFYLRKISDKGGFWELRVNESVVDNYLYGDLHIDNTRDFNKLNVK	1813	DR PROSITE; PS00118; SUBTILASE SER; UNKNOWN 1.
QY	481	DODIMPGMKYKANGRPDKYTMGDNVLYQGSDLNAKAVGHYQFLYKVKPEVNID	540	DR PROSITE; PS00678; WD REPEATS 1; UNKNOWN 1.
Db	1814	DODIMPGMKYKANGRPDKYTMGDNVLYQGSDLNAKAVGHYQFLYKVKPEVNID	1873	DR KW Cell wall; Peptidoglycan-anchor; Signal.
QY	541	PKGNTSLEYDGKSVVFNINDKNGDFGEQEHYINGKEYTSFDIKDILKLANK	600	FT SIGNAL_1 19 Potential.
Db	1874	PKGNTSLEYDGKSVVFNINDKNGDFGEQEHYINGKEYTSFDIKDILKLANK	1933	FT CHAIN 20 2144 PRTA. PRTA.
QY	601	IVVKDFARNTTKEFLINKDGTGEVSELKPHRTVTIONKEMSTIVSEEDFILPVKGE	660	FT SEQUENCE 2144 AA; 240725 MW; 205251470/41331 CRC64;
Db	1934	IVVKDFARNTTKEFLINKDGTGEVSELKPHRTVTIONKEMSTIVSEEDFILPVKGE	1993	DR Query Match 94.2%; Score 3793.5; DB 2; Length 2144; Best Local Similarity 94.7%; Pred. No. 1.3e-151; Indels 1; Gaps 1; Matches 732; Conservative 16; Mismatches 1; Gaps 1;
QY	661	LEKGYOPDGWEISGFECKDKAGYVINSKDKTIPKPFKKIEKEKKEENPKPTDFSKDN	720	DR 1 KLGBIASEKFKLNGNGKGSLSKDKTGGVHHHQENBSIKEKSSFTIDRNISTRFENK 60
Db	1994	LEKGYOPDGWEISGFECKDKAGYVINSKDKTIPKPFKKIEKEKKEENPKPTDFSKDN	2053	DR 61 DDKKLKRRKFREVDFFSETGRMRMERYDKDGGNITAYDGTOLYEYETKLDIEKSKI 120
QY	721	PQVNHSOLNESHRKEDLQREHSQSKDSTKOTATVLDKUNNISKSTNNPK	773	DR 1339 KLGBIASEKFKLNGNGKGSLSKDKTGGVHHHQENBSIKEKSSFTIDRNISTRFENK 1398
Db	2054	PQVNHSOLNESHRKEDLQREHSQSKDSTKOTATVLDKUNNISKSTNNPK	2106	DR 1399 DDKKLKRRKFQEDDFVN-GGTRTVERDYKYDGGNITAYDGTOLYEYETKLDIEKSKI 1457
RESULT 3				
Q9S4MB		PRELIMINARY; PRT; 2144 AA.		DR 121 YGLLSPSKDGHFELIGKISVNSKATYQGNYKSTEIKATKDFHSKTMFDLYANIND 180
ID	Q9S4MB			DR 1458 YGLLSPSKDGHFELIGKISVNSKATYQGNYKSTEIKATKDFHSKTMFDLYANIND 1517
AC	Q9S4MB			DR 181 IVGLABAGDMRLFVKUNDQKAEIKRMPKKETKSEPYVSSVNGAVELGEGDSLK 240
DT	01-MAY-2000	(T-EMBL; 13, Created)		DR 1518 IVGLAFAQDPLFVNDQKAEIKRMPKKETKSEPYVSSVNGAVELGEGDSLK 1577
DT	01-MAY-2000	(T-EMBL; 13, Last sequence update)		DR 241 KEDNLTMESGKYSDESKQYLKONTILRGYALKVTNPKGTDMLLEGNGVSKEDI 300
DT	01-MAR-2004	(T-EMBL; 26, Last annotation update)		DR 1578 KEDNLTMESGKYSDESKQYLKONTILRGYALKVTNPKGTDMLLEGNGVSKEDI 1637
DE		Cell wall-associated serine proteinase precursor PrtA.		DR 301 AKIOKAPNPRLASETTYIADSRSNVEGRSSTQVNSALDGFNIRYQVFPMKDGKA 360
GN		NameprtA;		DR 1638 AKIOKAPNPRLASETTYIADSRSNVEGRSSTQVNSALDGFNIRYQVFPMKDGKA 1697
OS		Streptococcus pneumoniae.		DR 361 IKDGNIVTDSKLVIFGKDKDKEYTGEDKNEVAKEKDGSMLIDTPVNLSMDKVFNP 420
OC		Bacteria; Firmicutes; Lactobacillales; Streptococaceae; Streptococcus.		DR 1698 IKDGNIVTDSKLVIFGKDKDKEYTGEDKNEVAKEKDGSMLIDTPVNLSMDKVFNP 1757
OX		NCBI_TaxId=1313;		DR 421 SKSNKTYVRNPFYLRKISDKGGFWELRVNESVVDNYLYGDLHIDNTRDFNKLNVK 480
RN		[1] SEQUENCE FROM N.A.		DR 1758 SKSNKTYVRNPFYLRKISDKGGFWELRVNESVVDNYLYGDLHIDNTRDFNKLNVK 1817
RP				DR 481 DODIMPGMKYKANGRPDKYTMGDNVLYQGSDLNAKAVGHYQFLYKVKPEVNID 540
RC				DR 1818 DODIMPGMKYKANGRPDKYTMGDNVLYQGSDLNAKAVGHYQFLYKVKPEVNID 1877
RX		MEDLINE=21585565; PubMed=11728722;		DR QY 541 PKGNTSLEYDGKSVVFNINDKNGDFGEQEHYINGKEYTSFDIKDILKLANK 600
RA		Bethe G., Nau R., Weilmer A., Hakenbeck R., Reinert R.R., Heinz H.P.,		DR 1878 PKGNTSLEYDGKSVVFNINDKNGDFGEQEHYINGKEYTSFDIKDILKLANK 1937
RA		Zybk G.; The cell wall-associated serine protease PrtA: a highly conserved		DR 601 IVVKDFARNTTKEFLINKDGTGEVSELKPHRTVTIONKEMSTIVSEEDFILPVKGE 660
RT		virulence factor of <i>Streptococcus pneumoniae</i> . ¹		DR 1938 IVVKDFARNTTKEFLINKDGTGEVSELKPHRTVTIONKEMSTIVSEEDFILPVKGE 1997
RL		FEMS Microbiol Lett. 205:99-104 (2001).		DR QY 661 LEKGYOPDGWEISGFECKDKAGYVINSKDKTIPKPFKKIEKEKKEENPKPTDFSKDN 720
CC		1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (BY SIMILARITY).		DR 1998 LEKGYOPDGWEISGFECKDKAGYVINSKDKTIPKPFKKIEKEKKEENPKPTDFSKDN 2057
CC		EMBL: AF127143; AAD68399.1; -.		DR QY 721 PQVNHSOLNESHRKEDLQREHSQSKDSTKOTATVLDKUNNISKSTNNPK 773
DR		HSSP; P00782; 2SBT.		DR 2058 PQVNHSOLNESHRKEDLQREHSQSKDSTKOTATVLDKUNNISKSTNNPK 2110
DR		MEROPS; S008.064; -.		
DR		GO: GO:0009986; C:cell surface; IEA.		
DR		GO; GO:00561; C:cell wall; IEA.		
DR		GO; GO:00823; F:peptidase activity; IEA.		
DR		GO: GO:0004289; F:subtilase activity; IEA.		
DR		GO; GO:000650; P:proteinase and peptidolysis; IEA.		
DR		Interpro; IPR01435; DUF1034.		
DR		Interpro; IPR01899; Gram_pos_anchor.		
DR		InterPro; IPR003137; PA.		
DR		InterPro; IPR000209; Pept_S8_553.		
DR		InterPro; IPR010259; Prot_inh_S8A.		
DR		InterPro; IPR01680; WD40.		
DR		Pfam; PF00680; DUF1034; 1.		
DR		Pfam; PF00746; Gram_pos_anchor; 1.		
DR		Pfam; PF00882; Peptidase_S8; 1.		
DR		Pfam; PF05922; Subtilisin_N; 1.		
RESULT 4				
AC	Q8DQP7	PRELIMINARY; PRT; 2144 AA.		
AC	Q8DQP7			

DT 01-MAR-2003 (TREMBIrel. 23, Created)
 DT 01-MAR-2003 (TREMBIrel. 23, Last sequence update)
 DT 01-MAR-2004 (TREMBIrel. 26, Last annotation update)
 DE Cell wall-associated serine proteinase *Prta* (EC 3.4.21.-).
 DE Name-prta, *Order* *Loccoccus*, *Name*=*pr561*,
 OS *Streptococcus pneumoniae* (strain ARCC BAA-255 / R6).
 OC *Bacteria*; *Firmicutes*; *Lactobacillales*; *Streptococcaceae*;
 OC *NCBI_TaxID*=171101;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21429245; PubMed=11544234;
 RX DOI=10.1128/JB.183.19.5709-5717.2001;
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszczak L.C., Burgett S.,
 DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
 Gilmour R., Glass J.S., Khoja H., Kraft A.R., Laiace R.E.,
 LeBlanc D.J., Lee L.N., Lefkojia E.J., Lu J., Matsushima P.,
 McAlister S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
 Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
 RA Sun P.-M., Winkler M.B., Yang Y., Young-Bellido M., Zhao G.,
 Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,
 RA Glass J.I.,
 RT "Glass of the bacterium *Streptococcus pneumoniae* strain R6";
 RL J. Bacteriol. 183:5709-5717(2001).
 CC [-] SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 an amide bond (By similarity).
 CC EMBL; AE008434; AAC09365.1; -.
 DR PIR: A97942; A37942.
 DR HSSP: P00782; 2SBT.
 DR MERO5; S88.064; -.
 DR GO: GO:000986; C:cell surface; IEA.
 DR GO: GO:004823; F:peptidase activity; IEA.
 DR GO: GO:004422; P:subtilase activity; IEA.
 DR GO: GO:0006508; P:proteolysis and Peptidolysis; IEA.
 DR InterPro: IPR010435; DUF1034.
 DR InterPro: IPR01899; Gram_pos_anchor.
 DR InterPro: IPR013137; PA.
 DR InterPro: IPR010259; Pept_inh_S8A.
 DR Pfam: PF00920; Peptidase S8; 1.
 DR Pfam: PF00920; Subtilisin; 1.
 DR Prints: PRO0723; SUBSILIN.
 DR TIGRFams: TIGR01167; LPXTG anchor; 1.
 DR PROSITE; PS00177; SUBTILASE_HIS; UNKNOWN 1.
 DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN 1.
 DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN 1.
 KW Cell wall; Complete proteome; Peptidoglycan-anchor.
 SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4BDBC503A0C CRC64;
 Query Match 92.1%; Score 3709.5; DB 2; Length 2144;
 Best Local Similarity 92.4%; Pred. No. 4.5e-148;
 Matches 714; Conservative 26, Mismatches 32; Indels 1; Gaps 1;
 Oy 1 KLGELAISKFKNGLGNGKGSKKDTGVEHHOENERSIKKSSFTIDRNSTIROPENK 60
 Db 1339 KLGELAISKFKNGLGNGKGSKKDTGVEHHOENERSIKKSSFTIDRNSTIROPENK 1397
 Db 61 DILKLLKKRKFRRVDFSETGKRMEEYDKDKGNTIAYDGTOLYEYETKDELSKI 120
 Oy 1398 DILKLLKKRKFRRVDFSETGKRMEEYDKDKGNTIAYDGTOLYEYETKDELSKI 1457
 Oy 121 YVILSFSKDRGHLIGKISNSKAKYVYGNKSLAKAKDPSKTFKMPDILANIND 180
 Db 1458 YVGLSFSKDRGHLIGKISNSKAKYVYGNKSLAKAKDPSKTFKMPDILANIND 1517
 Oy 181 YVGLAFLAGDMLFLVKONDQKAEIKTRMPKEKETKSEYVYVSSGNVRLBEGEDLSK 240
 Db 1518 YVGLAFLAGDMLFLVKONDQKAEIKTRMPKEKETKSEYVYVSSGNVRLBEGEDLSK 1577

QY 241 KPDNLTKMESKIKYDSKEQOYLKDNMLRKGYALKVTPNGKPTDMLLEGNGVYKED 300
 QY ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||| 1637
 Db 1578 KPDNLTKMESKIKYDSKEQOYLKDNMLRKGYALKVTPNGKPTDMLLEGNGVYKED 1637
 QY 301 AKIQKANPNURALSETTIVADSNTVEDGRSTQSVLMSALGFGNITYQVTFKNDKGEA 360
 QY ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||| 1697
 Db 1638 AKIQKANPNURALSETTIVADSNTVEDGRSTQSVLMSALGFGNITYQVTFKNDKGEA 1697
 QY ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||| 1757
 Db 1761 IDKGDMVTDSSKLVLFKGDKDKEYGDKFVNVEATKEDGSMFLFDTKPVNLMSMDKNGFNP 420
 QY ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||| 1757
 Db 1798 IDKGDMVTDSSKLVLFKGDKDKEYGDKFVNVEATKEDGSMFLFDTKPVNLMSMDKNGFNP 1757
 QY 421 SKSNKTYVRAPEFYLRGKISDKGGENWELRVNESVVDVNYLYGDLHIDNTRDENTKLNVK 480
 QY ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||| 1817
 Db 1758 SKSNKTYVRAPEFYLRGKISDKGGENWELRVNESVVDVNYLYGDLHIDNTRDENTKLNVK 1817
 QY 481 DGDIMDGKMDKYGKANGFPDKUTIDMCGWVYIQTGKSDLNAGKAVGHQFLYDNKVEWID 540
 QY ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||| 1877
 Db 1818 DGDIMDGKMDKYGKANGFPDKUTIDMCGWVYIQTGKSDLNAGKAVGHQFLYDNKVEWID 1877
 QY 541 PKGNTSIEYADGKSUVVINDKRNNGFDGETOEQHIVYINGKEYTSFNDIKQIDKTLN 600
 QY ||||||| ||||||| ||||||| ||||||| ||||||| ||| 1937
 Db 1878 PKGNTSIEYADGKSUVVINDKRNNGFDGETOEQHIVYINGKEYTSFNDIKQIDKTLN 1937
 QY 601 IWKQFARNNTYKEFLNKOTGEGVSELKPHRVTUIONGKMSSTIVSEEDFLPYVKG 660
 QY ||||||| ||||||| ||||||| ||||||| ||||||| ||| 1997
 Db 1938 IWKQFARNNTYKEFLNKOTGEGVSELKPHRVTUIONGKMSSTIVSEEDFLPYVKG 1997
 QY 661 LEKGQFDGWMISGEGKKGKAGYVWLNLSKOTPKIYKPFKEKEBENKTFDQKTDXILN 720
 Db 1998 LEKGQFDGWMISGEGKKGKAGYVWLNLSKOTPKIYKPFKEKEBENKTFDQKTDXILN 2057
 QY 721 PQVNISQLNESRKEDLQREBHSQSDSTKDTVATUDKONISSESTTNPNK 773
 Db 2058 PQVNISQLNESRKEDLQREBHSQSDSTKDTVATUDKONISSESTTNPNK 2110

RESULT 5

Q7RAS7 ID PRELIMINARY; PRT; 2649 AA.

AC Q7RAS7; ID PRELIMINARY; PRT; 2649 AA.

DT 01-MAR-2004 (TREMBIrel. 26, Created)
 DT 01-MAR-2004 (TREMBIrel. 26, Last sequence update)
 DE Hypothetical protein.
 GN Name=PY06422;
 OS Plasmodium yoelii yoelii;
 OC Bukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=32319;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XNL;
 RX PubMed=12368865; DOI=10.1038/nature01099;
 RA Carlson J.M., Antiguedad S.V., Sun B.B., Kool J.T.W., Peeta M.,
 RA Silva J.C., Ermakova M.D., Allen J.E., Selengut J.D., Koo H.L.,
 RA Peterson J.D., Pop M., Kobsack D.S., Shumway M.F., Bidwell S.L.,
 RA Shallom S.J., Van Aken S.E., Riedmiller S.B., Fieldblyum T.V.,
 RA Cho J.K., Quackenbush J., Seeger M., Shoaibi A., Cummings L.M.,
 RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdivia A.B.,
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 RA Carucci D.J.;
 RT "Genome sequence and comparative analysis of the model rodent malaria
 parasite *Plasmodium yoelii* yoelii";
 RL Nature 419:512-519(2002).
 CC -i- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC Preliminary data.
 DR InterPro: IPR011591; Botulinum.
 DR ProDom: PD001963; Botulinum; 2.

KW	Hypothetical protein.	Db	1933 KEELGKKNIS-----KDTINLEITTTKKKNPISKSNN 1966
SQ	SEQUENCE	SEQUENCE	SEQUENCE
Query Match	6.7%; Score 270.5; DB 2; Length 2649;	RESULT 6
Best Local Similarity	20.6%; Pred. No. 0.0052;	ID 081B84	PRELIMINARY;
Matches	206; Conservative 160; Mismatches 327; Indels 307; Gaps 49;	ID 081B84	PRT; 1642 AA.
QY	1 KUGBIABSKFKNLGN---GKEGSLKKDTGVEHHHOENESIKEKSFTIDRNISTIRD 56	AC 081B84;	
Db	1044 KONNLIERNYRNNDNLHQVNRAKONKNDISNLNSKSTGTYQNRSVITTEFIRS 1103	DT 01-MAR-2003 (TEmBLrel. 23, Created)	
QY	57 FR-----NKLKLUKLFKEPREVDDFTSETGKRMVEYDVKYDDKNTIAYDD----- 102	DT 01-MAR-2003 (TEmBLrel. 23, Last sequence update)	
Db	1104 FEDKIKIKLQNLQERLKKIEKUDEKEKVNDY-KKMEKIGKQD--NLFBATKHEK 1159	DE Hypothetical protein MAL8P1.29.	
QY	103 -GTDLEYETEKEDEIKSKTGYGLPSKDGHEFTLGKISNSVSKNAKUYYGN--NYKSTEIK 159	GN Name=MAL8P1.29;	
Db	1160 LACLEENMSKONBLIKNL-----KDKDMLKIELERSQIN-NYASNEHLANTLNUVK 1211	OS Plasmodium falciparum (isolate 3D7).	
QY	214 KETKSEPYVYSSYGNVILGEGDPLSKRNKPDLN----- 245	OC Bukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
Db	1272 KLLNMKGHNINYEYERSIDKLKSIDKLUKEQNLKLTQICDKEKNIMLEKAQLRNDESS 1331	RA [1] RA Seeger, K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,	
QY	246 -----TKEQSG-----KYS----- 256	RA Quail M., Barrell B., Harris D., Berriman M., Pain A., Hall N.,	
Db	1332 NTTTISIDSGTTINNEKIMMKEBEIAEALYKDKLKLKLNKLEEKTKNKLNTLTKTNEQS 1391	RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.	
QY	257 --SEKOOLKLONILAKGALKVTTYNPGKTDMLBEGNGVSKEDIAKIQKANPRL 313	DR EMBL; AL84507; CAD1123.1; ..	
Db	1392 ITFLANKKIKCLLKENKLNQKVNPKSI-----NDLKLNKTD-EDITKENSIPSI- 1442	KW Hypothetical protein.	
QY	314 SERTIYADSRNVEDGRSTQSVMSALD--GENIR--YQVTFKOMDKG-----EAD 362	SEQUENCE 1642 AA; 194753 MW; 701F5D28B2BB8097 CRC64;	
Db	1443 --TTFDLANKELEHNTVKQDTESTQIDSYDMARIKKGYDTSNHKONSENCYDQIKTEID 1500	Query Match 6.6%; Score 265.5; DB 2; Length 1642;	
QY	363 KO-----GNU--VTDSSKVLFGKDKPEY--GDKFNTAEIKDGSMLFIDTPKPV- 409	Best Local Similarity 20.4%; Pred. No. 0.0048; Matches 190; Conservative 133; Mismatches 313; Indels 297; Gaps 46;	
Db	1501 KQMYKEKCTPNUKETQDTDNLNTISVDKKEMTENQDYN-----IILEDNNLKNETNSMP 1557	Matches 190; Conservative 133; Mismatches 313; Indels 297; Gaps 46;	
QY	410 NLISMDKOVFNPSPSKNSKVNKFPEFY-----LRGKTSKCG---FNUWLRVNEWSVD 457	Db 104 ELNRNRNRRNRRNRRNRRNNSNSPIMDEKEKTNRNTMLKHKRKYDPPNTS-----DEK 158	
Db	1558 NNRKENTYTNLKKENKVNWNFEDSRIRANLQNNFESANGNESPNYEVNNNE--H 1614	QY 61 DILKLIKKEFREVDDFTSETGKRM---EVDYKDKGKNTIAVDDGTDLEYETEKLDE 115	
QY	458 NTY-----YGDHTDNTDRENK--LNVKDGDIMDWGMDKCYANGFPKVTMDGKV 508	Db 159 NKIR-----KKLNDINKYKKGMLPSSYENDEKYSKGSFNISLD-----DE 201	
Db	1615 NTNNMKINAYNSNHSLN---NIESEHVSQNTESLO-----NVNKVVKIYAPTEVD 1664	QY 176 ANINDIVDGLAFAGDMFLFKVNDQKAEKIRKMPKEKEKSEVYSSYGNVILGEG 235	
QY	509 YQOTGSDPLANKAVGVL-----YQFLYDNTKPEVNDP 541	Db 236 DKNN-----EMKEY-----STDSDIIRRVEISHSKE-----NNNDIKLKCH 275	
Db	1665 YHNN-----NLOPISLKHNDENRRTT-SENGTISYSNEKNDNSKTYMNNSKEYTEN 1719	QY 236 DISKNKDNLTKNSGKLYSDEKQOYLLKONILRKGYA----- 275	
QY	542 KGNTSIEYADGKSUVFNINDK-----RNGFGEIQFOHQIYINGKEYTSFNDIKOIDK 595	Db 276 NLIRNKQSNLVNFDRK-----RKENFIFLANYMKKGKENGNGKTNINSYDIEHDKDK 330	
Db	1720 NHRNSIKSD-----NWNKVKTECLNKINNEFANNESTLNNTETERNSTNDLNKIYE 1773	QY 276 -----LKVTTPNGK-----TDMLEG-----GVSKE-DIAKIQKANPRLS 314	
QY	596 TUNIKI-----VVKDFARNNTVKEFLN-RDTGEVSELKPHRVVTIONKEMNSITVS-EE 650	Db 331 NEILNNLDINSYNEKBLCSLYNNKENDFNKIGISKOPNVDSKERKYNKEKDHNLYNDS 390	
Db	1774 DNNIAYNNKIKIENY-NDODLKNYLNNSOKNNNNHEKND--NNTNEKKKKGEAWIDIKN 1831	QY 315 ETIYADSRNVEDGR-STQSVMSALDGFNTIRYQVFFKNDK----- 357	
QY	651 DFLPQVYKGELEKGYQFGWEGFEGKDDAGTVINLNSKDTFIPKPFVKKEKEEENKP 710	Db 391 DGKSKIRTNTVDDNSNSHNSLNL--ENFSNFMIDLEKDKRDKEEVKLIKSLRNN 448	
Db	1832 NEVLPYTK-----IECKVLSD-----EKEAGTKKRNK-----RKSCOKRNKST 1872	QY 358 -----GEATDKD-GNLVTD-----SSKVLFGKDDKEVGDKNFVNEIAKE 397	
QY	711 -----TPDVK-----KRDQFQVN-H-----SOLNESH 733	Db 449 EYENEKEGETHKKODBLNKVLVLDMTKIDNNEKKKKVLYKDYSE---NYERKISS 503	
Db	1873 VNNMRNRNTYIVRPSIESIKMKTEFANKKKNPNLHMKTNGKLDSLVNDISKLVNKI 1932	QY 398 DESMLFDTKPNLNSMKQYFNPSPSKNSKVNKFPEFYDVKYDGGFDDKVTMDGKV-----YL 510	
QY	734 KEDLQREERHSQKSDSTKQVT--ATVLDKNNISSKSTNN 770	Db 504 EN-----IDIGPTNRSFELVGDGNNTKDMWHINENDR-----DKNDN-----INN-----D 544	
QY		QY 511 QTCYSDJNAKAVGVLQFLYDNV-KPBNVNDPKGN-----TSIEYACKSUVFNINDKRN 565	
Db		Db 597 NSGNYNTDAN-----LSDQISNBANLKVNTNSMDYTKYDDK-----GNTKN 642	
QY		QY 566 GFDGEIQBOHTYINGKEYTSFNDIKOJIDKTLNKIVVKDFARNTTVREFLINKDQEVIS 625	

Db 643 TP---KKSAYDKNKKLIDTKLUMIGHNN---QKENTRND-EKVILNGINNSPK 691
 QY 626 ELPKPHRTVYTO---NGKEMSTIVTSEEDFTLPVYKGELEKGYQFDGWEISGFEGKED- 680
 Db 692 DADNKRSNIDLOVNNNNNNNNNNNTYDAMEAISGNDEQYKVNQED-HVIMGVENNN 750
 QY 681 -----AGYVINSKDTPI-----KPVF- 697
 Db 751 MQTNCSSNNKNYINNDRWNLKERYFDMMKPGUVUDNINTKENEVLQNEKKDIFNK 810
 QY 698 ---KKIEEKKEENK-PTFDVSKKKDNPNQVNH-----SOLNEH---RKELORE 740
 Db 811 SHGNEKEIPLDNNKNVKNVILKQVNDIHDYKVSNIGNEIDENRKKKENLNS 870
 QY 741 HSQKSDSTKDVATVLDKNNNISKSSTNNP 772
 Db 871 INQEKGIGNKVILINKTSYSKNNLSSYAEKH 903

RESULT 7

Q81L45 PRELIMINARY; PRT; 3504 AA.
 ID Q81L45; AC 081L45; DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DE Hypothetical protein.
 GN ORFName=Pf14_0404;
 OS Plasmodium falciparum (isolate 3D7).
 OC Bukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN NCBI_TaxID=3629;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2255705; PubMed=12368864; DOI=10.1038/nature01097;
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
 RA Carlton J.M., Pain A., Nelson K.B., Bowman S.G., Puisl T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
 RA Chan M.S., Nene V., Shallom S.J., Suh B., Peterson J., Angulo S.,
 RA Pertea M., Allen J., Selengut J., Haft D., Roos D.S., Ralph S.A.,
 RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.M., Barrell B.,
 RT "Genome sequence of the human malaria parasite Plasmodium
 falciparum,";
 RL Nature 419:498-511(2002).
 DR EMBL; AE01482; AAN7017.1; -.
 DR HSSP; Q93TC2; INWX.
 SQ SEQUENCE 3504 AA; 408308 MW; B9454D48D55BB4F0 CRC64;

Query Match 6.6%; Score 264; DB 2; Length 3504;
 Best Local Similarity 21.2%; Pred: No. 0.014;
 Matches 196; Conservative 131; Mismatches 346; Indels 252; Gaps 43;
 QY 11 KNLGNGKEGSKKDTGVEHHQEN-BESTKEK---SSFTIDRNISTIRDENKDL 62
 Db 1432 KNHENSEDNTIYTEEYCGKNSPQENTDDNTPKEKGNGKNSQQNSHNTWENGNKNS 1491
 QY 63 KKLKIKKPREVDDFTBTGKMEYVYKDK---GNTIAYDDGDLVEYETEKEBKS 118
 Db 1492 QQ-----NSDHNTIYTEEYCGKNSPQENTDDNTPKEKGNGKNSQQNSHNTWENGNKNS 1539
 QY 119 KIYGV---LSPSKD-GHFEETLGKISVNSVKAVVYG---NNYKSIERKATKDF--- 165
 Db 1540 -YEIDKRNRIHENDQHFTTYESRINKEFHNDVNTVSGDNNEKGKSVNNSEDIDP 1597
 QY 166 -HSKIMTFDLYANINDVGLAFAGDMRLFKNDOKKAERKIRMEPEKETKE--YP 221
 Db 1598 GKGKNIQDDILTEEN-----GKSKFENIEEKDLSKIQ-KSKISHEAEGHFTP 1647
 QY 222 VVS-----SYKVNVIELGEGDLSKPKDNLTMSKGK-----SDSEKOQYL 264

RESULT 8

Q7RGP2 PRELIMINARY; PRT; 1850 AA.
 ID Q7RGP2; AC Q7RGP2; DT 01-MAR-2004 (TREMBLrel. 26, Created)
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
 DE Hypothetical protein.
 GN Name=Pf04304;
 OS Plasmodium yoelii yoelii.
 OC Bukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN NCBI_TaxID=73239;
 RN [1]
 RP SEQUENCE FROM N.A.
 STRAIN=17XNL;
 RX PubMed=12368865; DOI=10.1038/nature01099;
 RA Carlton J.M., Angulo S.V., Suh B.B., Koo J.W., Pertea M.,
 RA Silva J.C., Brnoiaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.P., Biwell S.L.,
 RA Shallow S.J., van Aken S.E., Riedmiller S.B., Feldblum T.V.,
 RA Cho J.K., Quackenbush J., Seeger M., Shoaibi A., Cummings L.M.,
 RA Flinders L., Yates P.R. III, Raine J.D., Sinden R.E., Harris M.A.,
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,

RA	Carucci D.J.,
RT	"Genome sequence and comparative analysis of the model rodent malaria parasite <i>Plasmodium yoelii</i> yoelii."
RL	Nature 419:512-519 (2002).
CC	-1- CAUTION: The sequence shown here is derived from an preliminary data.
DR	EMBL; AAB101001295; EAA16146.1; -.
KW	Hypothetical protein.
SQ	SEQUENCE 1850 AA; 215313 MW; 00366ABD5CB21629 CRC64;
Query Match	6.5%; Score 260; DB 2; Length 1850; Matches 198; Conservative 154; Mismatches 379; Indels 298; Gaps 44;
Qy	6 AESKFKULGNG---KEGLSKKDTGVEHHRHOENESIKEKSSFTIDRNSTIR--- 55
Db	62 SIEKSYNDKONBKIGYKFNNSNSTKNTGKSTNSYDNTYADDYINE-----DNNAISFNSVCIN 115
Qy	56 ---DFEN-KDLKLUKKEF---RVDPFPTSETGKRMEEVDYK---YDDKGNITAYD---D 102
Db	116 SKLDYNSNLNRKSLKLEONFHNNNNDDDSNDSEBDYDGFKIRRYKCECSFNARDKID 175
Qy	103 GATDLEYETEKLBIKISKYGVISP-----SKQHFEI---- 134
Db	176 NAEKKYSLLEKNTIDN-IGPAPINSTSGDNILKRGKCTCNISEYNSMMNKEVNTA 233
Qy	135 --LGKISNVNSQNAKVYGNVYKSIKATKDFHSKTMF----- 172
Db	234 LENVKCENDKCNMSMEYKNDDVYKINTDSEINOREEENITKSPSTDIALSRISKDDDN 293
Qy	173 -----DLYANINDIVDGLAFLAGDM-RLFVKONDQKAEK----- 206
Db	294 LQDNKNDIYANNDMLNQNTMPOQDMKQVLTQDGNNINKPTKPSKSYFEVLYKETAGKNN 353
Qy	207 ---IRMPBKIKETKSEY ---PVYSSY ---GNTVIEGEGDLSKNEKDNLNTKESKLY 254
Db	354 ISBLETPLNLKODKEDYSGNTPIAYYLAGOPENKSLNGEEDMLENKISSFNLK-EIDKDY 412
Qy	255 SIEQKQVQLLKONILRG-----YALKVTTNPKGKTM-----EGNQYTSKEDIAK 302
Db	413 SKENFHSPFKDINTKEKGDEPPEABERTEBETENNIKONFSKRNHRHFSNNNIIID 472
Qy	303 IOKANPN-----LRALESTT-----IYADSRNVED-----GRSTOSVMSAL 339
Db	473 VNGQNPNUDKDELFIRRVSVISVKYKNIKTEBEYNMLKRDIKLSSNKSNSRKSTNN 532
Qy	340 DGENITRYQVFTEK-----MNDKGFAIDKDGNLVTDSSKLVLFGKDDKEYTGEDKFN 391
Db	533 NGDLILE-KVNTERKINNNDSVFTDKLIFSCNNNINMNTIFKDNLP--KEPKYSKLDINE 589
Qy	392 VEAIKEGQGSMMLF-----DTKPVNLSMDKMYFUPSKSKIKYVANPEFVLGKISDK 442
Db	590 INPEKYKGKDMLYSINDKNGINRENGLISKNTTFFNTENDYDQIDRITDPLNKLUM 649
Qy	443 GGNWELRNLVNESV-----DNYLTYGDLDHIDNTRDFNKLNVKGDM-----DWKGKDYKANG 496
Db	650 YGKSYDDITNGINNNNNNNXNNTYNNPNTLGSKPF-----IRSSWMINQWQFKGDINN 704
Qy	497 FPDKVTDMDGNYVLQGTGYSDSLNAKAVGPHYQFLDNYVKEPEVNTDPKGNTSIEVADGKSV 556
Db	705 IPINRSKSYTNTIGGNSNSDIYKNTNNINY-----NNKFMNSNNDPKSGYQVQDTEGENIN 759
Qy	557 FN-----INDKRG-----FDGBTQEOHQHINYGEYT 584
Db	760 PRNNSYEKKRNTEKYLINDKNAKREKNNYINXKMTTHDDNTFGDT-----INKRNTN 813
Qy	585 SEN-DIKOIKDNLNKLKV-----VQDFARNTVKEF---ILNKOTGEVSLK 628
Db	814 NFHIDNPNPLLNLINNINRNGNSVKGDNINKVENEONTEOFFKONDITNYSGGED-N 872
Qy	629 PHRVVTIQLNGKMSSTIVSEEDFILPVYK---GELEKGYQFDGEWEISGF--EGKODA-- 681
Db	873 NSKVNLIMEKSNTPNKSIISENDGFFPLINKVIANRLESTADP-GYKESPYMEKEDBN 931
Qy	682 -----GYNVIMSKD---TFKPKVFKIEKEENKPTFVSK 717
Db	932 KNFGIHTLGNDINNNNNENNNNGYNHDTKVKWVDFMKSFNDI-DVNEBKNLNDYKV 990
Qy	718 KNPQVNHSQLN-----EHSRKDQIQRREHSQKSDSTKDVATV-----D 758
Db	991 NNNSEKANKMDMEKKNSFYKDPNEDTLAKYIHEQKGSITEVVEKILSHKEINFINGE 1050
Qy	759 KNNISST 767
Db	1051 KONIKUGT 1059
RESULT 9	
QYRR9	ID 07RR9 PRELIMINARY: PRT: 2757 AA.
AC	07RR9;
DT	01-MAR-2004 (TREMBrel. 26, Last sequence update)
DT	01-MAR-2004 (TREMBrel. 26, Last annotation update)
DE	Rhoptry protein.
GN	Name=PT00649;
OS	Plasmodium yoelii yoelii.
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX	NCBI_TaxID:7329; [1]
RN	SEQUENCE FROM N.A.
RC	STRAIN=17XNL;
RX	PubMed=12368865, DOI=10.1038/nature01099;
RA	Carlton J.M., Angiuolo S.V., Suh B.B., Kooij T.W., Pertea M., Saliva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L., Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L., Shallom S.J., van Aken S.B., Riedmiller S.B., Feldblyum T.V., Cho J.K., Quackenbush J.J., Sedgwick M., Shoaibi A., Cummings L.M., Florens L., Yates F.R., III, Raine J.D., Sinden R.E., Harris M.A., Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B., Smith H.O., White O.R., van Lin L.H., Janse C.J., Waters A.P., Smith M.J., Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J., Carucci D.J.,
RT	"Genome sequence and comparative analysis of the model rodent malaria parasite <i>Plasmodium yoelii</i> yoelii."
RL	Nature 419:512-519 (2002).
CC	-1- CAUTION: The sequence shown here is derived from an preliminary data.
DR	EMBL; AAB101001298; EAA1709.1; -.
DR	InterPro; IPR001623; DnaJ_N.
DR	InterPro; IPR00499; ReticulocyteBP.
DR	TIGR4M; TIGR0612; 235kDa-fam; 1.
DR	PROSITE; PS00635; DNAl_1; UNKNOWN_1.
DR	SEQUENCE: 2757 AA; 324332 MW; 55AD73DB2B9EFA37 CRC64;
Query Match	6.4%; Score 259; DB 2; Length 2757; Best Local Similarity 20.6%; Pred. No. 0.017; Mismatches 326; Indels 276; Gaps 43; Matches 198; Conservative 163; Mismatches 326; Indels 276; Gaps 43;
Qy	18 ESSLKKTGTTGVEH-----HOENESIKEKSSP-----TIDRN----- 51
Db	542 EAGVKKSYEVKWKRKSLKLEBENENKVKLETQIKLNFNOVYKIKIDBNVYKLNKL 601
Qy	52 ---STRDFENK-----DLKLUKKEFREVDDFTSETGKRMEEY----- 87
Db	602 LELKEKJKNISKDKNEYKKAVDILKKEIENNAVYDLSAKTSQVQFVRLKSDTIVSTIK 661
Qy	88 -----DVKDQGKNTIAYDDCTDELEYE-----KLDENKSKVYKVLSPRSKDG 130
Db	662 SELSQIVFDDDKLYNLISLVSQVQENDNVEDDKLQDLSQKIDNVSKIQMETAVES 721
Qy	131 HEBIL---GKISNVNSQNAKVY-GYNVYKSLIKAKMDFHSKTMFDPYANINDWGL 185
Db	722 HLNTIETNKNLSDTIVEKIKYIGBISK--DLNKTEPDKNEK--ELSNKIND---- 772

Qy	730	ESRKEDIQREHHSOKSDSTKVTAT-VLDKNNISSKTTNNP	771	Oy	285	KUDMLEGGVYVSE- DIAKTOKANPNRALSETTYADSRNEDGRSTOSVILMSALD-G	341
Db	1188	DSNDSSD-----SNKSRSQGQIKTKKGSNSKKECP	1224	Db	765	KKLIEDRKSBEETYKINTLKVBDYKVHT-- NELISNFRNKQILKOKDKN	819
				Db	342	ENIIRY----QVFTK--- MNDKGEAIDKGNLY----- TDSSKLVIF----- G	378
RESULT 11				Db	820	INTVKENNSIDKLYTKEPFENLTDKKEFLDKFTDVALNDHESNNNEMKYFNNLKKANLG	879
Q7RMS4		PRELIMINARY;	PRT; 2661 AA.	Qy	379	KDKD---EVGDEDKNEAETKEDGSMMLFDTKPVNLSMOKYFVPSKSWKLYVNPFFY	434
ID	Q7RMS4;			Db	880	LRGKISDKGGFNWELRVNESVUDVLYGDIHIDNTRDPNPKLNKQGDIMDGWMKDYKA	939
DT	01-MAR-2004	(TREMBLrel. 26, Created)		Qy	435	KKKKNMYYQQGEBEKRIDDKKKNAIDENVSKIEVASYIMISDEMEREIKIES	977
DT	01-MAR-2004	(TREMBLrel. 26, Last sequence update)		Db	940	LNTQVIEK----- VKTNTNLAKIGKUKQVDPFGKGENIK-----	977
DE	Rhoptry protein (Fragment).			Qy	495	NCFPDKYTDMDGCVVYHQTYGSDLNAKAVGVVHOFUDNPKVEVNIDPKGNTSIEADGK	554
GN	Name=Py02104;			Db	978	--YPDEINKKNEI----- KVVG-- QDQDHINKLKEIKNNSXNDEIKGKT	1021
OS	Plasmodium yoelii yoelii.			Qy	555	VVF-NNDKRNPGDCEIQEHHYVINGKEYTSFN-DIKQIDTKTNIKIVKDPARTTV	612
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			Db	1022	DKFENTDKE----- TYNKDPKEFEKIQI----- VTKDEKNIYI	1059
RN				Qy	613	KEFILNDGTGEVSELKEHRT----- VTLONGKEMSTIVSEDFILPVVKGELEKGYQ	666
RX	PubMed=1268865; DOI=10.1038/nature01099;			Db	1060	K--INKLDSIEBKNTSLKKVCDINSYGOSIGKPLEQID-----	1101
RA	Carlton J.M., Angiuoli S.V., Suh B.B., Koo H.L.,			Qy	667	FDGWEGSGFESKDKDAGVYVNLNSKDTPIKPV- FKKLEKEGEENKPTFD----- VSKKDN	720
RA	Peterson J.D., Pop M., Kosack D.S., Shumway M.F.,			Db	1102	--EKKKAGHMK-SMEYVMDLNDLKKKSQETENEMKINMDKMYKINKEA	1151
RA	Riedmiller S.B., Feldblyum T.V.,			Qy	721	PQVNHSQNLNEHRSKEDIQREHHSOKSDSTKVTATWLDKNN--SSKSTNNP	772
RA	Florens L., Yates P.R. III, Raine J.D., Sinden R.E.,			Db	1152	LNISHDDQPH--NISKKHEBKISD----- IHRNSLKIQEFSTESNIN	1194
RA	van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,						
RA	Carucci D.J.,						
RT	"Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii.";						
RL	Nature 419:512-519(2002).						
CC	- CAUTION: The sequence shown here is derived from an						
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is						
CC	preliminary data.						
CC	EMBL; AAB0100391; EAA0802-1; -.						
CC	InterPro; IPR011591; Botulinum; 1.						
Db	708 AMKNNELLRKKVIFDFDGAEEKTTELVKKIRIEKLVNYYERK-- FNNS						
Db	764						

SQ	SEQUENCE	Match	Score	DB	Length	Matches
Best Local Similarity	1389 AA;	6.4%;	Score 257;	DB 2;	Length 1389;	
Local Similarity	193;	Conservative	20.4%;	Pred. No.	0.0089;	Mismatches
						288;
					Index	326;
					Gaps	48;
QY	KEGSLKKKOTTGVEHHHQE-----					
Db	354 KKEAFQKEKNSSERLHERNIPYKEKEILEASKYDVTYSYFNKEEIKDKGSNNYDNKGVIPs	57 FENKLKKLKKERKEVDDFTSEKRMEEYDVKDKGKNTIAVDGDTADLYESEKLBRI	116	NEESIKEKSFTIDRNISTRD	56	
QY		57 FENKLKKLKKERKEVDDFTSEKRMEEYDVKDKGKNTIAVDGDTADLYESEKLBRI	116			
Db		414 EINKNVQRINKCNKCNHVSHSDT-----EKYE-----SNTS-----DLYNQNDK	457			
QY		414 EINKNVQRINKCNKCNHVSHSDT-----EKYE-----SNTS-----DLYNQNDK	457			
QY		117 KSKSIVGVLSPSKDGHFELIGKISWNSKNAKVTGN-NYKSIKAT-----KY	163			
Db		458 KLNIGYL--SKEDY--SNDKQKQDTEEVYDNEKDPTDQVNDKVLDSPLTHY	510			
QY		458 KLNIGYL--SKEDY--SNDKQKQDTEEVYDNEKDPTDQVNDKVLDSPLTHY	510			
Db		164 DFHISK-TWTFDLYANINDIVDGLAFLAGDMRLFVKNDOKKAEKIR--MPEKIKETKSB	219			
QY		164 DFHISK-TWTFDLYANINDIVDGLAFLAGDMRLFVKNDOKKAEKIR--MPEKIKETKSB	219			
Db		511 KYSNVILINKNIKNIK-----EINNITDVQVQKPGMENVILPHK-----556	556			
QY		511 KYSNVILINKNIKNIK-----EINNITDVQVQKPGMENVILPHK-----556	556			
Db		220 YPTVSYGAVI-----ELGEGDLSKRNPDMLTOMESKGKYSDEK-----QYI	263			
QY		220 YPTVSYGAVI-----ELGEGDLSKRNPDMLTOMESKGKYSDEK-----QYI	263			
Db		557 -----YSNDILKIOIKTULSEENENDIKTINKE--TONEATSYQOSDRNSTOPNSDRN	607			
QY		557 -----YSNDILKIOIKTULSEENENDIKTINKE--TONEATSYQOSDRNSTOPNSDRN	607			
QY		264 LKONITILRKGYALKV-----TYPNGCPTDMLLEGNGVSKED- 299	299			
Db		608 IKONIMDRKTCISKCYVYLPLANNINHEINKOLLINISTYE-----MEKEKL	657			
QY		608 IKONIMDRKTCISKCYVYLPLANNINHEINKOLLINISTYE-----MEKEKL	657			
Db		300 -----IAKLOKANPMLRALSETTIYADSRNVEDRGRSTOSVY-MRAGDEN---TIRY-O-V	349			
QY		300 -----IAKLOKANPMLRALSETTIYADSRNVEDRGRSTOSVY-MRAGDEN---TIRY-O-V	349			
Db		658 LPDTIYLIQ--SKSISSEKYMISIKCKVNGYLYPILNLENVNLNDENKNPWPINSONNN	715			
QY		658 LPDTIYLIQ--SKSISSEKYMISIKCKVNGYLYPILNLENVNLNDENKNPWPINSONNN	715			
Db		350 FTRK-----MNKGAEADKD-GMUVTDSKUJLFGKODKEYCDEKPNVAKTEKDGSM	402			
QY		350 FTRK-----MNKGAEADKD-GMUVTDSKUJLFGKODKEYCDEKPNVAKTEKDGSM	402			
Db		716 STPKPTVYSLYNE-----IDREIGNL-NYKNU-----NKUKIKLUPKGKYY	759			
QY		716 STPKPTVYSLYNE-----IDREIGNL-NYKNU-----NKUKIKLUPKGKYY	759			
Db		403 -----FIDTKEVNLSMDK-----YFNSKSNKLY	427			
QY		760 INDICPRFFDKNNVLUVCUDNNRKNSGNGNQIGONLIEDHIGHVETANYVFLNMSNFI	819			
Db		428 VRNP-----EFLRGKISDKKGFFNMLRVNESWVNLYLIGDHLTN	469			
QY		428 VRNP-----EFLRGKISDKKGFFNMLRVNESWVNLYLIGDHLTN	469			
Db		820 VKSSLDDLFFSHIPLFPLIEQVYFDFTEKNELLKQKFTMFLAITEACINICFL-----872	872			
QY		820 VKSSLDDLFFSHIPLFPLIEQVYFDFTEKNELLKQKFTMFLAITEACINICFL-----872	872			
Db		470 TRDENIKLNUKVGDDIMDGWMDKYGANGFDDKVTMDGNYLQGYSDSLANKAVGWHQEL	529			
QY		470 TRDENIKLNUKVGDDIMDGWMDKYGANGFDDKVTMDGNYLQGYSDSLANKAVGWHQEL	529			
Db		873 ---FNIAGI-----DSLFLY--KNLQAQVQTKDENDNPFPLTYSSE-RVQNGKHCYF	923			
QY		873 ---FNIAGI-----DSLFLY--KNLQAQVQTKDENDNPFPLTYSSE-RVQNGKHCYF	923			
Db		530 YDNVTRP-----EVNTDPKGNTSIEVAD-GKSVVFNINDKRNNGFDEGLOEBOHYYI	578			
QY		530 YDNVTRP-----EVNTDPKGNTSIEVAD-GKSVVFNINDKRNNGFDEGLOEBOHYYI	578			
Db		924 LNTYSKPLIFRPOSENELN-----NISHADVEGHINLNTSDERHKCEEEIKKKEIE	976			
QY		924 LNTYSKPLIFRPOSENELN-----NISHADVEGHINLNTSDERHKCEEEIKKKEIE	976			
Db		579 NGKEYTSENDIQLQIKDUNIKIWKDFARNTTKEFINKDGTGEVSELSKPHRTVTLN	638			
QY		579 NGKEYTSENDIQLQIKDUNIKIWKDFARNTTKEFINKDGTGEVSELSKPHRTVTLN	638			
Db		977 NG-----IK-LIKKTNNDFFTYDARYQKSI--FLN-----AIKNDDAKNTEN	1018			
QY		977 NG-----IK-LIKKTNNDFFTYDARYQKSI--FLN-----AIKNDDAKNTEN	1018			
Db		639 GKEMNSTIVSBEDE-----ILPYVGELEK-----GQF-----DGW	670			
QY		639 GKEMNSTIVSBEDE-----ILPYVGELEK-----GQF-----DGW	670			
Db		1019 NYDKRQKQIYTHSYTQKQFQDYYINILNPNQOLQKHKEDYSFRREVETLQSNRQG	1078			
QY		1019 NYDKRQKQIYTHSYTQKQFQDYYINILNPNQOLQKHKEDYSFRREVETLQSNRQG	1078			
Db		671 BISGEGKEDAG-YVNLISKDFTPKVFKIEKEE-----ENKPTFDVSKCNDNPQVN	724			
QY		671 BISGEGKEDAG-YVNLISKDFTPKVFKIEKEE-----ENKPTFDVSKCNDNPQVN	724			
Db		725 HSQLNESHRKEDLQREBHSQKS DSTKDVATVUDKNNSSKSTIN	770			
QY		725 HSQLNESHRKEDLQREBHSQKS DSTKDVATVUDKNNSSKSTIN	770			
Db		1132 KOTPSDKKEPONITYDENTONLN-----NNFSNTQNN	1165			

ID 0813P4 PRELIMINARY; PRT; 1777 AA.
 AC 0813P4;
 DT 01-MAR-2003 (TREMBrel; 23, Created)
 DT 01-MAR-2003 (TREMBrel; 23, Last sequence update)
 DT 01-MAR-2003 (TREMBrel; 23, Last annotation update)
 DB Hypothetical protein PFE1095w;
 GN Name=PFE1095w;
 OC Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OC NCBI TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks D.,
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
 RA Cronin A., Davies R., Daviss P., Dear P., Dearnard F., Doggett J.,
 RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
 RA Humphray S., Jadesk K., James K.D., Johnson D., Kerhoro A.,
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Leonard N.,
 RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
 RA Oliver K., Ormond D., Price C., Quail M.A., Robbins-Bostach E.,
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
 RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.;"
 RL Nature 419:527-531(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Devlin K., Baker S., Davies P., Mungall K., Berriman M., Pain A.,
 RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.,
 DR Submitted (2002) to the EMBL/GenBank/DBJ databases.
 DR InterPro: IPR011591; Botulinum.
 DR Product: PDB01963; Botulinum; 1.
 DR Hypothetical protein.
 KW SEQUENCE 1777 AA; 213320 MW; 244467CFF190C522 CRC64;
 SQ

Query Match 6.3%; Score 255; DB 2; Length 1777;
 Best Local Similarity 20.7%; Pred. No. 0.015; DB 2;
 Matches 186; Conservative 152; Mismatches 324; Indels 236; Gaps 44;
 Mismatches 186; Conservative 152; Mismatches 324; Indels 236; Gaps 44;
 QY 4 EIAESKFQNGLNGKEGSLLKDKTGVHHHQNE-----ESIKKSSFTIDRNNTIRDF 57
 QY 58 ENK--DLKLLI-----KKKFREUDDFTSETGKRMEEVYKDDKGNNIAYDGT 104
 Db 1031 EENLLDDKKKKDDEENELLDKDKKKLDEENILUBERKKKDEONILLDEKKKEITNDHNT 1090
 QY 977 ELLDDKKKKL--DBENEELLDKDK--KKKDDEENELLDKCKKKLDEENELLDKDKKKL--DE 1030

QY 105 DLEYETEKLEIKSKI--YGVLSPS-KDGHF----EILGKISVNSKNAK----- 146
 Db 1091 FQOTEHNLINNEKTKIQDYNIEAELKEMHSKKLLEETKIKENDENDIKRKQSQIEN 1150

QY 147 VYGNNTYKSIETKATKDYKDFHSKUTMFYDLYANIDNDVGLAFAGDMRLFVKNDGKAEIK 206
 Db 1151 IYRNNSMMIDNINSYSYKSNVKTIFNFNNIE----- 1184

QY 207 IRMPEKIKETKSEYPYVSSYGNVIELGEGDLSKNPKPDWLTKMGSKVSDSEKQYLUKD 266
 Db 1185 --EDKTOKONTKITYTDISNM-----LTKNNKSSYISINSKEDITINEOEIYSNKN 1234

QY 267 NI-----ILRKYALKVTTYNFGTKTDMLEGNGVSK----EDIAKIQK-ANPN 309
 Db 1235 NTFNNDIEKDLTNUVMERKONNNNNISYKCCQSNTYDINSNNKFLMDTESIKIQDINEK 1294

QY 310 LRLSL--ETTIVADSRVNTEDGRSTOSVUM-SALDGFNTIYQVFPTKMDKGEADLKDGN 366
 Db 1295 VVUJKQREBFVKEFKOFKEKEVNLHLKJENVJMKNNIKDR--EENLQKKEEPLKMK 1352

RA Fraser C.M.; Barrell B.;
 RT Genome sequence of the human malaria parasite Plasmodium
 RT falciparum.;
 RL Nature 419:498-511(2002).
 DR EML: AE014841; AANZ5938.1; -.
 DR HSSP; Q93IC2; IMWI.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 227 AA; 262841 MW; 6D5D4B8BF643339C CRC64;
 Query Match 6,3%; Score 252.5; DB 2; Length 227;
 Best Local Similarity 19.1%; Pred. No. 0.024;
 Matches 190; Conservative 155; Mismatches 319; Indels 329; Gaps 49;
 QY 10 FKNLGNCKE-----GSLKQKTTGQH-----HHGENESEIKEK 42
 Db 690 YNNNNNNKETCTSINIKHSELKYPFNKSHVNSYMKVNTNHLPHRNATTSNNNNNEYEKEK 749
 QY 43 SSFTIDRNIST---IRDFFNK---DLKLU-----KKFREVDFT-----SE 79
 Db 750 EK---DRNITNGNNNTVEYVNNSCIPPLKOMIPDGVRNHSINKKUNVNTQRTSSVY 806
 QY 80 TCKRMEEYDY-----KYDKKNTIAYDDGTLESTEKDEIKSISIYGVLSPSK 128
 Db 807 TNKNDIDENSPDMPIINGIRESKVSKVNNNT--NGSIGFFSKLQHNYHQSNNV----- 858
 QY 129 DGHFETLGKISNVSKNAVKVYVGNVYKSEIEMATKVDPHSKMTP-----DLYANIN----- 179
 Db 859 -----NESYPLKNNMMKNN-----IEHNDKKNIFLVKNVSYEDTYSNTHNGI 900
 QY 180 -----DVIDGLPAGDDPR-----LFVUDN-----DQK----- 202
 Db 901 HENSMILKVNLYKKACTPHGYSRHNOKMYTHERNTNOKKSYHYHNGTTLKPLVNTNN 960
 QY 203 -----ABIKRMPKEKIKETKS---EVYVASYGNVIELGEGDLSKPKDMLTKMSEGI 253
 Db 961 WVNNEFADINLSAQKLUHSLSMGYEDKSMENYRNKLYNNNNNNNN-----NI 1013
 QY 254 YSDSEKOOY---LLKONTILRGYAL-----KVTTYNPGKTDMLBEGNGVYSKEDTAK--- 302
 Db 1014 YNDNEYCQYNNSYCFDHSQDLMNPFLPHQNSKULTHSNNNKNSFFNGINVESSHHLANPEI 1073
 QY 303 -----IQKANP----- 309
 Db 1074 KTFARHNSYPILNQGLINCPLQCLGDSNORHKNHVVYIKHNEYLNNIGSIINVLRKG 1133
 QY 310 LRALS-----ETTYADSRNEDGRSTOSVLSMALDGNIIRYQVTFKND 356
 Db 1134 LAKTISTHNGKESFSNMDNKVYMEGUNIQONVANNNNKESC-DN1KHMTRKSLAFVSRE 1192
 QY 357 K-GE---AIDKQDNLVUTDSSKUVLFGK-DDKCYTGEDKFVNVAIKED-GSMMPID---TKP 408
 Db 1193 SYGEHKSLSLDVQBCYVKGKLI--NKVNDKKT--EDNNNSYLNEDDNASMQFYETNSNP 1248
 QY 409 VNLSMD---KVNFPNSKSNKTVRNBFYRKGKISKGGEW-----LRNEST----- 455
 Db 1249 YTVDOENNNMKYVN---NVLVNNNSNVYUDSKVNDKSKENAEKNSDDILNNENIHTKD 1304
 QY 456 -----VDNLYLTYGDLAHDNTRDFNIKVNVDGDDIMW--GMDYKANGFPDKV-TDMG 506
 Db 1305 QDKKIQNNNEFESEQADIENTRSQSEVYEHPEI-WVTRASNEEKSKTELYSDMSS 1363
 QY 507 NYLQTCYSDIMAKAVCHYQFLYDNVTPENIDPKNTSTSAYADCKSVVFNINDRNNG 566
 Db 1364 NRVTKNKYSDDMNVNEVNLNEONLTLTEKVKYQLEKENKMDMVTVEENINTIKTEND 1423
 QY 567 PFGEIQRQHITYINGKEYTSFEDIKQIDKTKNTKTVKDFPRT--TVK--EFLANKDTG 622
 Db 1424 INEEVR-----NEQRESINHIND-----TMNINHITDEYNDTYPNIKOTECVHNENN 1472
 QY 623 EVSELKEPHRVTTIQNGKEMSTTIVSBDPFLPVKGEBEGYQFFGWEISGPEGKKGAD 682
 Db 1473 MNSIBQYTFPHTRNNHLVK---NNQNP-----FEEBGLNEENFECK--V 1516

QY 683 YVINISKD-----TPIKPVFKCIEKKEBENKPF-----DYSKKDN 720
 Db 1517 YIENNKKDDKGDSKTSNLTSRNICKSSENDHNEKENTIVVRKGEGKIKRKVSMKGRN 1576
 QY 721 PQVN-HSOLN-----ESRKEDLOREHSOKS 746
 Db 1577 EKLNEBENYNNYDQMDNRHQNDDTREKENDEN 1609

Search completed: February 16, 2005, 19:19:52
 Job time : 188 secs

Db 1694 IDKGDNILVTDSSKLVLFGKDKDKEYTSDKENVATEAKEDGSMFLIDTPKPVLNSMDKNYFNP 1753
 Qy 421 SKSNKTYVRNPFYLRGKISDKGGFWELRNBNSVDNYLTYGDLHIDNTRDPNKLNV 480
 Db 1754 SKSNKTYVRNPFYLRGKISDKGGFWELRNBNSVDNYLTYGDLHIDNTRDPNKLNV 1813
 Qy 481 DGDIMWGMKDKYKANGFPDKTIDMDGAVYQFLTGYSIDLNAKAVGVAQVHQLFLDNVKPEVNID 540
 Db 1814 DGDIMWGMKDKYKANGFPDKTIDMDGAVYQFLTGYSIDLNAKAVGVAQVHQLFLDNVKPEVNID 1873
 Qy 541 PKGNTSIEYAGKSVVFNINDKRNNGFGEQEHVYINGKEMSTIVSERDILPVYKE 600
 Db 1874 PKGNTSIEYADGKSVVFNINDKRNNGFGEQEHVYINGKEMSTIVSERDILPVYKE 1933
 Qy 601 IWKDOPARNTVKEFLNKOTGEVSELKPRHTVTINGKEMSTIVSERDILPVYKE 660
 Db 1934 IWKDOPARNTVKEFLNKOTGEVSELKPRHTVTINGKEMSTIVSERDILPVYKE 1993
 Qy 661 LEKGKOPDGWISGFRKKDQAGYVINSKOTPIKPKKLERKSEBNKPTFDVSKKDN 720
 Db 1994 LEKGKOPDGWISGFRKKDQAGYVINSKOTPIKPKKLERKSEBNKPTFDVSKKDN 2053
 Qy 721 PONVHSQNLNSHRSKEDLQREHRSOKSDSTKWTATVLDKNNISKSTNNPK 773
 Db 2054 PONVHSQNLNSHRSKEDLQREHRSOKSDSTKWTATVLDKNNISKSTNNPK 2106

RESULT 2

A97942 metalloproteinase (EC 3.4.21.-) A [imported] - *Streptococcus pneumoniae* (strain R6)
 C;Species: *Streptococcus pneumoniae*
 C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
 C;Accession: A97942
 R;Huskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; E;e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAnren, S.; M;Y, P.M.; Winkler, M.E.
 J; Bacteriol. 183, 5709-5717, 2001
 A;Author: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A;Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.
 A;Reference number: A97872; MUID:21429245; PMID:1544239
 A;Accession: A97942
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-2144 <KUR>
 A;Cross-references: UNIPROT:Q8DQP7; GB:AE007317; PIDN:AAK99365.1; PID:915458138; GSPDB:C
 C;Genetics: DRTA
 C;Keywords: hydrolase; serine proteinase

Query Match 92.1%; Score 3709.5; DB 2; Length 2144;
 Best Local Similarity 92.4%; Pred. No. 6.4e-151;
 Matches 714; Conservative 26; Mismatches 32; Indels 1; Gaps 1;

Qy 1 KLGELTAESKFPNLNGKEGSKKDTGVEHHQHOBESIKESKSTIDRNISTRD芬K 60
 Db 1339 KLGELTAESKFPNLNGKEGSKKDTGVEHHQHOBESIKESKSTIDRNISTRD芬K 1397
 Qy 61 DDKLKKKKPKEVDDTSETGKMERDYKDKGMIAYDDGTLVYERKDKI 120
 Db 1398 DDKLKKKKPKEVDDTSETGKMERDYKDKGMIAYDDGTLVYERKDKI 1457
 Qy 121 YGVLSPSKDKGHEILGKISNSKNAKVVYGNYKSEIKAQKTDPSKTMFLYANIND 180
 Db 1458 YGVLSPSKDKGHEILGKISNSKNAKVVYGNYKSEIKAQKTDPSKTMFLYANIND 1517
 Qy 181 YDGLAFAFGDKRFLVNDOKRAEKIRMPKKEKTKSEYFVYSSYGNVIEFLGDSK 240
 Db 1518 YDGLAFAFGDKRFLVNDQKRAETKIRMPKKEKTKSEYFVYSSYGNVIEFLGDSK 1577
 Qy 241 KPDNLTKMSEKSYSEKQYLLKOMILRGKALKVTTNPKGTDMLLEGNTYSKED 300
 Db 1578 KPDNLTKMSEKSYSEKQYLLKOMILRGKALKVTTNPKGTDMLLEGNTYSKED 1637

Qy 301 AKIQKANPRLASETTIVADSRNTEDGRSTOSVMSALGFTNRYQVTPKMDKGEA 360
 Db 1638 AKIQKANPRLVLSSETTIVADSRNTEDGRSTQAVLMSALGFTNRYQVTPKMDKGEA 1697
 Db 361 IDKGDNILVTDSSKLVLFGKDKDKEYTSDKENVATEAKEDGSMFLIDTPKPVLNSMDKNYFNP 1757
 Qy 421 SKSNKTYVRNPFYLRGKISDKGGFWELRNBNSVDNYLTYGDLHIDNTRDPNKLNV 480
 Db 1758 SKSNKTYVRNPFYLRGKISDKGGFWELRNBNSVDNYLTYGDLHIDNTRDPNKLNV 1817
 Qy 481 DGDIMWGMKDKYKANGFPDKTIDMDGAVYQFLTGYSIDLNAKAVGVAQVHQLFLDNVKPEVNID 540
 Db 1818 DGDIMWGMKDKYKANGFPDKTIDMDGAVYQFLTGYSIDLNAKAVGVAQVHQLFLDNVKPEVNID 1877
 Qy 541 PKGNTSIEYADGKSVVFNINDKRNNGFGEQEHVYINGKEMSTIVSERDILPVYKE 600
 Db 1878 PKGNTSIEYADGKSVVFNINDKRNNGFGEQEHVYINGKEMSTIVSERDILPVYKE 1937
 Qy 601 IWKDOPARNTVKEFLNKOTGEVSELKPRHTVTINGKEMSTIVSERDILPVYKE 660
 Db 1938 IWKDOPARNTVKEFLNKOTGEVSELKPRHTVTINGKEMSTIVSERDILPVYKE 1997
 Qy 661 LEKGKOPDGWISGFRKKDQAGYVINSKOTPIKPKKLERKSEBNKPTFDVSKKDN 720
 Db 1998 LEKGKOPDGWISGFRKKDQAGYVINSKOTPIKPKKLERKSEBNKPTFDVSKKDN 2057

RESULT 3

T28317 ORP MSV156 hypothetical protein - *Melanoplus sanguinipes* entomopoxvirus
 C;Species: *Melanoplus sanguinipes* entomopoxvirus
 C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
 C;Accession: T28317
 R;Atonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, B.; Kutish, G.F.; Rock, D.L.
 J; Virol. 73, 533-552, 1999
 A;Title: The genome of *Melanoplus sanguinipes* entomopoxvirus.
 A;Reference number: Z20484; MUID:9102612; PMID:9847359
 A;Accession: T28317
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-1127 <AFO>
 A;Cross-references: UNIPROT:O9YV16; EMBL:AF063866; NID:94049647; PIDN:AC97677.1; PID:94
 C;Genetics: C;Keywords: C;Note: MSV156

Query Match 6.3%; Score 253.5; DB 2; Length 1127;
 Best Local Similarity 21.5%; Pred. No. 0.0015;
 Matches 186; Conservative 155; Mismatches 338; Indels 185; Gaps 44;

Qy 7 ESKFKNLNGKEGSKKDTGVEHHQHOBESIKESKSTIDRNISTRD芬K 63
 Db 196 EIEFKNDNQKEKTKSEYFVYSSYGNVIEFLGDSK 255
 Qy 64 KLIKKKPKEVDDTSETGKMERDYKDKGMIAYDDGTLVYERKDKI 111
 Db 256 FNIDERQKLDQINSKINTLNENTIKGVMNLTYETPNKLNQNETNKOStKSIDEQK 315
 Qy 112 KLDERKSKYGVLSPSKDKGHEILGKISNSKNAKVVYGNYKSEIKAQKTDPSKTMFLYANIND 161
 Db 316 LLDDELDKNTNNTSFLYNSK---TKTNTIQOLQLESSLTDENNANI---NINELSKK 367
 Qy 162 KYDFHSKTMFLYANINDYDGLAFAFGDKRFLVYERKDKI 216
 Db 368 LFNDIQKLNDITEQNQKTD--FFNNSTRIKEKLDTEYKKDIDKINNQKLEESK 425
 Qy 217 ---KSEY---PYVSSYGNVIEFLGDSKDNKPNPLKONESGKYSDDSKQOYQVILKDNII 269

RESULT 7

T28676

rhoptry protein - Plasmodium yoelii (fragment)

C.Species: Plasmodium yoelii

C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C.Accession: T28676; A45521

R.Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.

Mol. Biochem. Parasitol. 76, 329-332, 1996

A.Title: Comparison of two members of a multi-gene family coding for high-molecular mass A.Reference number: 22057; MUID:91077455; PMID:920022

A.Accession: T28676

A.Status: preliminary; translated from GB/EMBL/DDBJ

A.Molecule type: DNA

A.Residues: 1-2401 <SIN>

A.Cross-references: UNIPROT:Q26216; EMBL:U36927; NID:91041784; PID:91041785; PIDN:AAB412

R.Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.

Mol. Biochem. Parasitol. 42, 241-246, 1990

A.Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple cDNA clones and their expression in Escherichia coli. A.Reference number: A45521; MUID:91101660; PMID:2270106

A.Accession: A45521

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 2260-401 <KRE>

A.Cross-references: GB:M34281

Query Match 5 7%; Score 230; DB 2; Length 2401;

Best Local Similarity 21.0%; Pred. No. 0.04; Gaps 46;

Matches 187; Conservative 138; Mismatches 292; Indels 274; Gaps 46;

2 LGEIMAEKSKPKLGNQKSGEKLKDTTGE-HHQQENESIKKSSFTIDRNISTIDPENK 60

589 IDKITYDKEFNLITLDTKTELETKPTGLSLNNHESNN--KELITYFDLKLNLGKONEN- 644

61 DLKLUJJKKKREVDDFTSETGKRMFEDYDKIDDKGMIA-----YDDGTDLIYET 110

645 ---MLYKQFNEKE- 111 EK-----LDEBKSKIVYLSPSKDGHFBLIGKLSNSKRNAAVYGN 151

694 GKSIELNLNTKULEKYGAVNTNLNTEKEK- 694 GKSIELNLNTKULEKYGAVNTNLNTEKEK- KAEDEIKKQVNDKINIVSNTBITIVTSIVNINEDNEI 693

RESULT 8
 T18440
 hypothetical protein C0425w - malaria parasite (*Plasmodium falciparum*)
C.Species: *Plasmodium falciparum*
C.Accession: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C. Lawson, D.; Bowman, S.; Barrell, B.
 submitted to the EMBL Data Library, August 1997
A.Reference number: Z18935
A.Accession: T18440
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-450 ^{LW>}
A.Cross-references: UNIPROT:O77336; EMBL:Z98547; NID:e1325376; PID:e1325396; PIC
C.Genetics:
C. Map position: 3
A.Note: C0425w

Query Match 5.7% Score 228; DB 2; Length 4550;
Best Local Similarity 21.6%; Pred. No. 0.11; Gaps 43;
Matches 193; Conservative 119; Mismatches 347; Indels 236; Gaps 43;

5 IAEISKFKNLNGKE--GSLKDDTGYV---EHHHQENERSIKERGSSFTIDRNISTIRD 56

A;Reference number: A24594; MUID:86014355; PMID:2995820

A;Accession: A24594

A;Molecule type: DNA

A;Residues: 1-1640 <HOL>

A;Cross-references: UNIPROT:P04933

C;Superfamily: major merozoite surface antigen

C;Keywords: surface antigen

Query Match 5.4%; Score 219; DB 2; Length 1640;

Best Local Similarity 18.6%; Pred. No. 0.071; Mismatches 340; Indels 228; Gaps 39; Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

Qy 23 KPTG-VERHHQNEESIK-----EKSFTDRNISIRDBNPKLKKFREVD 75

Db 232 KONGVKHEDYIKKNTTENLIEESKKTIDKNNKATKER-----KKLYQAOY 283

Qy 76 FTTSETGKRMEEVDYKDDKGNTIAYDDGTDLLEVETEKEUDKSKYIGVLSPSKDGHFEL 135

Db 284 DLSIYNIKOLEE-----AHLNLSV-----LE---KRIDLK-----KNNENKELL 319

Qy 136 GKLNVSKNAKVYGGNNYKSIETIKATKVDHF-----SKWTFDLYANINDIVGLAFA 188

Db 320 DKLNEKNNPPANSNTNTLIDKNNKIEBHEKEIETAKTKFNIDSFTDPL----- 373

Qy 189 GOMRLFVYKDNDQGKAEKIRMPPEKIKERKSEP-----YVSYGNV-----IELG-E-GDLK 239

Db 374 ELEYYIREKNN-KNTDISAKVETKESTEPEVNGVTPLSNDINNALNBLNSFGDLI- 430

Qy 240 NPKDNLTKMBSGKITYSDSEKKQYL-LKONIL-----RKGYALKVTTNPKGTDML 289

Db 431 -NPFDYTKEPSSKNIYTDNERKKINKEKEKKEKKESDKSYEYRSKLN-----BIT 485

Qy 290 EG-----NGVYSKE-----DIAKIOKA-----NPNRLASETTYADSRNVEDGRSTQSV 334

Db 486 KEYEKLINEIYDLSKFNNDLNTFEPKOMGKRVSYKEVLTHTPNTFASYENSKHNLKTK 545

Qy 335 LMSALDGE-----NII-----RQVFTKOMDKGEAIDKGNTIYDSSKLUVFGKDKEYT 385

Db 546 ALKYMEDSLRNLTVKEKLYKYYKLISKINENETLVE--NIKKDEBOLF-----EKKIT 598

Qy 386 GEDKFNVEAIKEGDSMLFDTKVALNSMDKMYNPKNSKNSKIVRNPEFYLRGKISDKGG 445

Db 599 KDNKPDPEKTLVEYDPIVKVQVKYL--MKNDKDELKKTQTLKVN-----LKHINHVN 653

Qy 446 NWELRVNNEHSVVDNYLTYKDPHTNTDRNENKL-----NVK----- 480

Db 654 KOENKQE-----PYYLIVLKKEIDKLKVFMKPVESLINEBEKKNKTQGOSDNPSTEGI 709

Qy 481 -----DSDIMDGMKDYK----- 501

Db 710 TQGATTKPGQOAGSALEGDSVQAQAEQKQAOQPPVPUVPEAKAQVPTPPAPVNTKENV 769

Qy 502 TMDGDNVYLQTGYSIDLNA-----KAVGHYQFLYDNVKPBNIDPKGNTSIEYADGKSVW 556

Db 770 SKL-----YLEKLYQFLNTSYICKYIUVSHSTMNEKLUKQYKIKERBSSKSCDPL 826.

Qy 557 FNIN-----DKRNUFGD-----EIQBQHIIYNGKEYTSFNDIKOIKDTNLKIV 603

Db 827 FNIQONNIPVMSMEDSLANLSLQFMEYKEVNCVNLKQNDKDKIKRNLLEAKKVTSV 886

Qy 604 KDPARNTTKEFLANKDGTGEVSEKPKHPRVVTIYONGKMSSTVSEEDFPLVYKGSLK 663

Db 887 KTLSSSSMOPSLSTPDKPKEVSANDDTSHSTNLANSLKLENTIS-----LGKNN 937

Qy 654 GYQPDGWEISGFSGK-----KDAGVNLNSKOTPKPVFKKEBKEEENKPTF-----D 713

Db 938 IYQ-----BLIGQKSENPKVKEKILSDFTYNESTPNFTKS KADDINSLANSERKLUED 993

Qy 714 VSKEKQDNPOVNSOLNESHRKEDIQREEHQSNSDSTKVTATVLKNNISK-STNNP 771

Db 994 INKLKTIQLSFLDLYNKYKLKLERLFDKIKTIVKQKOMIKULIKEOLESKLNSLNP 1052

RESULT 11

S4QKL

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st

C;Species: Plasmodium falciparum

C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jun-2000

C;Accession: A25120

R;Mackay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.; B

EMBO J. 4, 3823-3829, 1985

A;Title: Polymorphism of the precursor for the major surface antigens of Plasmodium falc

A;Reference number: A91030; MUID:86136024; PMID:3004972

A;Molecule type: DNA

A;Residues 1-1631 <MAC>

C;Comment: The merozoite stages of different strains have strain-specific surface antigens

C;Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The me

C;Superfamily: major merozoite surface antigen

C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane prote

F;67-84/Region: 3-residue repeats (S-G-TP)

F;1614-1631/Region: membrane anchor #status predicted <MBN>

F;97,259,759,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (Asn) (

Query Match 5.4%; Score 218; DB 1; length 1631;

Best Local Similarity 18.1%; Pred. No. 0.078; Matches 164; Conservative 162; Mismatches 339; Indels 240; Gaps 37; Matches 167; Conservative 164; Mismatches 339; Indels 240; Gaps 37;

Qy 23 KPTG-VERHHQNEESIK-----EKSFTDRNISIRDBNPKLKKFREVD 75

Db 232 KONGVKHEDYIKKNTTENLIEESKKTIDKNNKATKER-----KKLYQAOY 283

Qy 70 FREVDITSETGKRMEEVDYKDDKGNTIAYDDGTDLLEVETEKEUDKSKYIGVLSPSKD 129

Db 283 LSEAHNLISVLUKRID-----TILKNNENK----- 307

Qy 130 GHEFLGKISNYSKNAKVYGGNNYKSIETIKATKVDHF-----SKWTFDLYANINDIV 182

Db 308 ---BLDKINEKNNPPANSNTNTLIDKNNKIEBHEKEIETAKTKFNIDSFTDPL 364

Qy 183 DGLAPAGDMRLTVKDNQKQKAKIKRMEKPEKIKERKKEKKEKKEKKEKSYEDRSKSLN- 473

Db 365 -----ELEYTREKNN-KNTDISAKVETKESTEPEVNGVTPYPLSYNDINNALNBLNS 416

Qy 235 -GDLISKPKDNLTKMEEGKITSDEKKQYL-LKONIL-----RKGYALKVTTNP 283

Db 417 FGDLI--NPFDYTKEPSSKNIYTDNERKKFNEBKEKKEKKEKKEKSYEDRSKSLN- 473

Qy 284 GKDMLIG-----NGVYSKE-----DIAKIOKA-----NPNRLASETTYADSRNVEDG 328

Db 474 ---DTEKEYEKLINEIYDLSKFNNDLNTFEPKOMGKRVSYKEVLTHTNFASENKH 530

Qy 329 RSTQSVMSALDGP-----NII-----RQVFTKOMDKGEAIDKGNTIYDSSKLUVFGK 379

Db 531 LEKLTALKYMDLSRNLTVKEKLYKYYKLISKINENETLVE--NIKKDEBOLF----- 584

Qy 380 DDKEYTSEDKFNVEAIKEGDSMLFDTKPVNLSDMDKMYFNPNSKNSKIVRNPEFYLRGK 439

Db 585 -EKKITDKDENKDEKTLVEYDPIVKVQVKYL--MKNDKDELKKTQTLKVN-----LKHNI 638

Qy 440 SDKGFFWELRVNESVVDNYLTYGDLIUDNTDFNKL-----NVK----- 480

Db 639 HVPSYKQENKOB-----PYYLIVLKKEIDKLKVFMKPVESLINEEKKNKTQGQSDN 694

Qy 481 -----DGDIDMGMKDY----- 495

Db 695 STPGBTITQATTKPGQOAGSALEGDSVQAQOBKQOPPVPUVPEAKAQVPTPPAPV 754

Qy 496 GPPDKVMDGMYLQYQGSDLNA-----KAVGHYQFLYDNVKPENIDPKGNTSIEYA 550

Db 755 NKYENVKLID-----YLEKLYQFLNTSYICKYIUVSHSTMNEKLUKQKITKEBESKSSC 811

Qy 551 DGKSVVNN-----DKRNGEDG-----BIOQHIIYNGKEYTSFNDIKQIDKTL 597

Db 812 DPLDLLEUNIQNIPVPMWSMFSQNLNSVSQLMELIEKEMVNLYKIKNDKIKNLLEAK 871
 Qy 598 NIKIVVKDFARNVTVKERFLNKDTGVESELKPHRVVTIQLQNGKEMSTIVSEEDFLPVY 657
 Db 872 KVTSVKTLSLSSMQPLSLTPDKPEVSANDTSHSTNLNLNSKLHENILS----- 922
 Qy 658 KGBLEKQYQDFWEISPEGK-----KQAGYVNLSKOTPIKUVKKFEEKEENK 709
 Db 923 LGRKNKNYQ----ELIGQKSENFYEKILKQSDTFSNFSFTNPKVSKADDINSLDESKR 978
 Qy 710 PPF--DVKKQNPQVHNSQLNESHREKEDLQREHQSQKSDTSDKDVATVLDKNTSSK-S 766
 Db 979 KKEBEDINKLKLQLSFDLNNYKYLRLERFDKKKTVGKVQVQKLLKEQESKLN 1038
 Qy 767 TTINNP 771
 Db 1039 SLNNP 1043

RESULT 12

S05603 major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (strain 3D7) #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
 C;Species: Plasmodium falciparum
 C;Date: 12-Feb-1993
 C;Accession: S05603; S04550
 R;Myler, P.J.
 submitted to the EMBL Data Library, April 1989
 A;Reference number: S05603
 A;Accession: S05603
 A;Molecule type: mRNA
 A;Residues: 1-639 <MNY>
 A;Cross-references: UNIPROT:P04933; EMBL:X15063; NID:99896; PIDN:CAA33163.1; PID:99897
 Nucleic Acids Res. 17, 5401, 1989
 A;Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Plasmodium falciparum
 A;Reference number: S04850; MUID:8934516; PMID:266888/
 A;Accession: S04850
 A;Molecule type: mRNA
 A;Residues: 1504-1639 <MW12>
 A;Cross-references: EMBL:X15063
 C;Superfamily: major merozoite surface antigen
 C;Keywords: glycoprotein; merozoite; surface antigen
 C;P1-19/Domain: signal sequence #status predicted <SIG>
 C;P20-1639/Domain: major merozoite surface antigen #status predicted <MAT>
 Query Match 5.4%; Score 218; DB 2; Length 1639;
 Best Local Similarity 18.6%; Pred. No. 0.079;
 Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;
 Qy 23 KOTTG-VERHHOENESIK----EKSPTIDRNTSTIRDPENKOLKKLKKPFREDD 75
 Db 232 KONVKGMDYIKKKNTTENTNELLEBSKKTIDKNKNTKRE-----LGKLYQAOY 283
 Qy 76 FISETGKMEEDYKIDKGKNAIYDGTIDLEYETKDELSKSKIVGLSPSKDGFIL 135
 Db 284 DLSIYKQLE-----AHLNLSV----LG--KRDTK-----KNNENELL 319
 Qy 136 GKSNSVSKNAVYGNVYKSTBIKATKDFH-----SKTMFDLYANINDIVGFLA 188
 Db 320 DKINETKOPPNSGNTPNLTDKNNKIEERKEIKAIIKTFKIFNDSLFDPL----- 373
 Qy 189 GIMRLFVTKDNDQKKAETKIRHPEKIKETKSEV----YVSYGNV----IEGE-GDLSK 239
 Db 374 ELEYVYIREKON-KNIDISAKVSTKESTEPEVNPYNGVTPYPLSYNDKNNALNLNSFGDLI- 430
 Qy 240 WPDNLTKMEESSKJYSEKQYL-LKDNIL-----RGKALKVTTNGPKDML 289
 Db 431 -NPFDYTKPESPKNTYDNERKPKINEKEKIEKIEKIESKKSVDKSLN---DIT 485
 Qy 290 EG----NGVVSKE----DTAKIKA----NPNLRALSHPTVYDASRNVEDGRSTOSV 334
 Db 486 KEYEKULNEIDSKFNNIDLINFERNOMGKRYKSYKVERKLHNTASYENSKHNLK 545

RESULT 13

Q90603 Lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)
 C;Species: Mycoplasma pulmonis
 C;Accession: G90603
 R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.; Nucleic Acids Res. 29, 2145-2153, 2001
 A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis
 A;Reference number: A99512; MUID:2126715; PMID:11353084
 A;Accession: G90603
 A;Status: Preliminary
 A;Molecule type: DNA
 A;Residues: 1-1546 <KUR>
 A;Cross-references: UNIPROT:Q98P17; GB:AL445566; PID:914090150; PIDN:CA13908.1; GSPDB:G
 A;Experimental source: strain UAB CTIP
 C;Genetic code: SGC3
 A;Gene: MTFU_7350
 Query Match 5.4%; Score 217.5; DB 2; Length 1546;
 Best Local Similarity 20.2%; Pred. No. 0.077; Mismatches 336; Indels 249; Gaps 40;
 Matches 180; Conservative 125; Mismatches 336; Indels 249; Gaps 40;
 Qy 1 KLGELAEKSKFLGNGKQEGSILK----KDTGVERHHQHQEBSIEKSFTRDNTISTIRD 56
 Db 103 KSKELAESQVGTSVNNEBKFNLKFLRANKTKNTKPKSASKSOLNTSDQHLLNSKQSNISN 162
 Qy 57 FENKULKLKFKFREDDFTSETGKMEEDYKID-KD-KDNILAIYDGTIDLEYETKDE 115
 Db 163 IKNEQSKLQLTANNKEEHSANLQTONILESENTQODLDEK-----NSNKEKELKN 214
 Qy 116 IKSKYGVVSPSKDCHFETLGK-TSNVSKNAVYGNVYKSTBIKATKDFHSKMTFDL 174
 Db 215 VLS--FNKQEAVKQGSPEFKFKEVSKLKESKINIDKVDIEFLKADP----SFEQ 267

QY	2	LGEIMESKFKP---NLGNGKEBSSLKQ---KOTGVEHHHQBEEBILG
Db	266	LSEIKYKDKCTEISNSKRG---KDKIEFLERKFKPI
QY	52	-----STIRDD
Db	323	LKDIDBAEKQASTKVLFLHKHBTTSNIFKESSEBILG
QY	79	BTGKMRMEYDYKDKGNTIAYDDGTDLEYETEKLDD
Db	383	BIQTQKGQFQENLNKUNEPHYDNEAED-ELNNDKSTI
QY	139	SNVSKRKAQVYQVYNNYKSI-EIKATKQDFHSKUM-----
Db	438	TNIKOCGEGKLY-SKAQDIMOKIKATSSENTAEKLEKT
QY	191	MRLFLVKOND-----OKKAELKIRMMI
Db	491	RNLNTYTEKNRNLGIDSOPTINTEGALKESKGNGTEIGFF
QY	219	EYPYSSYGNVTELGEGDLSKRNKPDLNTKMEGSKTYY
Db	551	TVGNFISSSLFNNFDLNQYDFDNKNNINDYENKM---GFIYH
QY	279	TTYNQGKTDMLQEGNSYVSKEDI---AKIQKANPRLR
Db	606	SDYNSAKTRLSEA---OKEKVNLLNKEEANKYLR
QY	336	MSALDGFNIIRYQVTFPKMDKGAEADKDGNIYTDSS
Db	647	-----TIFNMK---BSUDKINEMIKEEKE
QY	396	KEDGSMFLFIPTKPVNLS-----MDKN-----
Db	684	KE-----LVDEN--NLSDILQOATGKNEEIQKITHST
QY	440	SDKGFFNWEILRVNESSVD-----NLLI
Db	737	TP-----ELALTELLGDAKLUKTAOBLSKFESKNNVW
QY	482	-----GDTMDW3MKDYKANGFPDKVTDMDGNYW
Db	791	VALETLAHSBIDTKOKD---SSKLIEMQNOIYV
QY	526	YQFLYDNVK-----PEVNIDPKGNISIEYADGKSVW
Db	845	VKI---GIVSKKXHSELSKITSCTSDSKSYDNTIALEKQTEI
QY	575	HYIYNGKEYTSFNDIKQIIDKTLNIKI-----
Db	901	-----KIKTDPESTNLN-KTLEGUNVALKASSDNH
QY	606	-FARNNTVKERPLNKTQDGEYSELKPHRHTVT---IONG
Db	954	IDSLMTALDELLKKGRCTEVSRYKLKDITVKEISD
QY	663	KGYQDGWEISGPEGKDKAGDVYVINSLSKDTF1KV
Db	745	KSDSPDVTA---TVD---KNNISSKSTN 770
QY	1010	KNYR-----DWTOD---VLTLINEHFTKQVSNFT
QY	714	--VSCKK-----DNPQVN
Db	1058	TISLKLKGVIEVNENTEMNTIESAKIEALYNELM
Db	1118	NADKUVIDVSKFNTDITOKSNTVNTQHNSINN 11495

Query Match 5.4%; Score 215.5; DB 2; Length 1252;
 Best Local Similarity 20.9%; Pred. No. 0.071;
 Matches 207; Conservative 130; Mismatches 324; Indels 331;

RESULT 15

C:Date: 22-Apr-1993 **#sequence_revision** 18-Nov-1994 **#text_change** 09-Jun-2000
C:Accession: A45597 **#sequence_revision** 18-Nov-1994 **#text_change** 09-Jun-2000
R:Li, W.B.; Bzik, D.J.; Tanaka, M.; Gu, H.M.; Fox, B.A.; Inselburg, J.
Mol. Biochem. Parasitol. 45: 229-239, 1991
A:Title: Characterization of the gene encoding the largest subunit of Plasmodium falciparum ribosomal RNA polymerase II
A:Accession: A45597; **MUID:**92018020; **PMID:**1656254
A:Status: preliminary
A:Residues: 1-2339 <u1>
A:Note: sequence extracted from NCBI backbone (NCBIN:61099, NCBI:61101)
C:Keywords: nucleotidyltransferase; transcription

Query Match	Score	DB	Length	Gaps
Best Local Similarity	5.3%;	214;	2;	2339;
Matched	21.6%;	Pred. No.	0 18;	Indels
191;	Conservative	134;	306;	254;
Mismatches	50;			
QY	24	DTRGVEHHHQE--NEESIKEKSFTIDRNISTIRD--ENKDLKLKKEPREVNDPTS	78	
Qy	1292	DNTVTEOLETMKELSKNKTKEKOSPK---GTIRDMHDSEBQMNKEITKAK---FPI	1342	
Qy	79	ETGKRMEEVDYKDDKGNTIAYDGGDTLEYETEKLDETKSKY-----G	122	
Db	1343	EK-KKGKHMBCNDIEVNTQD--NIQNNISCNVIKSQLENTHQVNNDLSIKN	1398	
Qy	123	VNSPKSKDCH-----FEILGKISNTSKNAKTYGQNNYKSTEIKAKYDFHSKTMFDL	174	
Db	1399	VILPKKEYHSIFHFRVNDYRNVVEKINLMDKCKFLNNSEK--WVOSKYNRMKLNKKI	1456	
Qy	175	YANINDIVD-----LAFAGDMLPKNDQ--KKAERKIRMEK---IKETKSE	219	
Qy	1457	EI--INNNYNEKEKLLNRKTMDDNTWSSDSDTAKKIKIKKEKRRKVKPKEKEN	1515	
Qy	220	YPVVSYGNVIELGEGLSKNPKEDNLTK-MESGKIVDSEKQOYLKDNTILRGYALKV	278	
Db	1516	FDR-NYKMTDNNDNN	1569	
Qy	279	TTYNPKGKTDML--EGNSVYKEDIAKIQKANPDL-----RALSETTYNDS--	322	
Db	1570	TNYN--TNYTYPNDNGIYEKETNNNELLTNNSNCNDKNDFSDEFFNNINENDLTLNKY	1626	
Qy	323	-----RNVEDGRSTQSOSVILMSALDGFNNIIRYQF--TPKMDKGEALDKOGNLVTPSSKL	374	
Db	1627	YRQIFKVNIGFVSVFVEVYESKOHYILPFYETIKWISFILEVLTETIPNTNLFHTLSK-	1685	
Qy	375	VLRQFKD--DKEYTGEDKVNKEKEDCSMFLDTKPVNLSMDKNTVNPSENKIVVRNP	431	
Db	1686	-----KEKTHOKTGKOMKQYIEIJK--WLF--RAINI--YKTFSKFIEL-----1728		
Db	432	EFYLRGKTSQDKGGRMELRVNESWVDMVLYGDLHIDNTRD----FNI---KLNVKD	481	
Db	1729	-----IKKKDYFNYIYIK-NYDISHRVIYH-DYSFINKOLYLPITFENIYKFKVISTP	1779	
Qy	482	GDIMDGWMDKDYKANGFPDKVTDMDGNGVNLQTGVSDDMKAVALAVSYHYFLYDNVKEPNIDP	541	
Db	1780	GD--AVGOSISAQSIGPCTQMTLKTHFAGVASHMV-TLGV----PRIKEINASN	1828	
Qy	542	KGNTSIEVADGKSVVFNINDKRUNGF-----DGETOBOHYINGKEYS-----F	586	
Db	1829	SICPPI-----LNPILLENDYNAFLAMMKSKELETTIRIDCIVK-EDYTRGVFESVKF	1882	
Qy	587	ND-----IKQIDKTLN-----KI--WVKDPAARTVK--BPIMLKD	621	
Db	1883	NEELIQKLFNLNINAYNITKDLKQSHINKIINKIHNIVNKYKLHISLKDNEF-----1938		
Qy	622	GEVSELKKEHRVVTITQNGKEMSTTIVSEEDFLPVYKGEBEKGYQFDGWEISGPFER-K	679	
Db	1939	FOMESLKKGKLDLILYGDQDKECIIKKEDI-----EVTDNEDEICD	1980	
Qy	680	DAGVUNISKOTFKPVVKIEKK--EEENKPTFVSKKDNPQVNHSQLNESIRKED	736	
Db	1981	DMRBYVYNSQG-----ELYERKCNKESNENKARVKKEID-----DN	2019	